

OM of: US-09-303-518d-571 to: SwissProt_40.* out_format : pfs

Date: Jun 30, 2002 8:44 AM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+np.model -DEV=xih
-O=/cgn2_1/USPRO.spool/US09303518/runat_28062002_142714_4362/app.query.fasta_1.23501
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=isp -GAPOP=12.000
-GAPXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=6.000 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomscore2
-TRX=human40.cdi -LIST=100 -DOCALL=200 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OTEM=pts
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USBR=US09303518.GCGNL.1.440 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XIPXY -WAIT -THREADS=1

Search information block:

Query: US-09-303-518d-571
Query length: 894
Database: SwissProt_40.*
Database sequences: 105224
Search time (sec): 217.960000

Score list:

| Sequence | Strd Orig | ZScore | EScore | Len | Documentation |
|---------------------------|-----------|--------|---------|------|---------------------------------|
| SwissProt_40:IDG_ECOLI + | 196.50 | 272.97 | 5.4e-08 | 306 | P24187 escherichia coli, ddd |
| SwissProt_40:MSB_HAEN + | 168.00 | 232.26 | 9.9e-06 | 318 | P76522 haemophilus coli, ddd |
| SwissProt_40:MSB_HAEN + | 139.00 | 191.37 | 0.0019 | 311 | P45567 haemophilus influenza |
| SwissProt_40:SGS_DROE + | 126.50 | 173.20 | 0.0182 | 328 | P13730 haemophilus influenza |
| SwissProt_40:YGR_HAEN + | 125.00 | 171.21 | 0.0238 | 323 | P24205 escherichia coli, ddd |
| SwissProt_40:MSB_HAEN + | 122.50 | 165.04 | 0.0369 | 437 | P13561 haemophilus coli, ddd |
| SwissProt_40:MSB_HAEN + | 120.50 | 167.75 | 0.0519 | 231 | P06659 haloflex sp. (strat) |
| SwissProt_40:HRK_HUMAN + | 119.00 | 152.02 | 0.0818 | 1104 | P01074 homo sapiens (human) |
| SwissProt_40:DBA_RAT + | 117.50 | 154.01 | 0.1017 | 688 | P10074 homo sapiens (human) |
| SwissProt_40:VE2_HPV1 + | 110.50 | 149.70 | 0.3365 | 361 | P06764 ratius norvegicus (rat) |
| SwissProt_40:MEP_LOJER + | 110.50 | 146.99 | 0.3489 | 493 | P26766 ratius norvegicus (rat) |
| SwissProt_40:YGR_HAEN + | 108.50 | 148.25 | 0.4752 | 308 | P040240 totium perenne (peten) |
| SwissProt_40:VE2_HPV1 + | 108.50 | 147.92 | 0.4773 | 320 | P33445 bordetella pertussis |
| SwissProt_40:GRTS_WHEAT + | 107.00 | 143.22 | 0.5512 | 506 | P22420 human papillomavirus t |
| SwissProt_40:GRTS_WHEAT + | 107.00 | 137.41 | 0.7009 | 839 | P10388 trititum aestivum (wt) |
| SwissProt_40:CCNA_RABIT + | 105.50 | 146.89 | 0.7886 | 221 | P062093 mus musculus (mouse) |
| SwissProt_40:SON_HUMAN + | 105.50 | 126.05 | 1.04 | 2424 | P27884 oryctolagus cuniculu |
| SwissProt_40:YGR_HAEN + | 105.00 | 126.05 | 1.04 | 2426 | P18583 homo sapiens (human) |
| SwissProt_40:CFE_CHLVC + | 105.00 | 142.41 | 0.9081 | 253 | P18824 aplysia calliorhiza |
| SwissProt_40:555_RAT + | 105.00 | 134.72 | 1.01 | 825 | P063003 ratius norvegicus (rat) |
| SwissProt_40:REP_MOUSE + | 105.00 | 131.98 | 1.04 | 1130 | P07347 mus musculus (mouse) |
| SwissProt_40:YV4_STRER + | 104.50 | 140.42 | 1.01 | 395 | P20188 streptomyces fradiae |
| SwissProt_40:SON_MOUSE + | 103.50 | 138.13 | 1.04 | 514 | P06921 human papillomavirus t |
| SwissProt_40:VE2_HPV1 + | 102.50 | 142.64 | 1.36 | 2404 | P09947 mus musculus (mouse) |
| SwissProt_40:HRK_HUMAN + | 102.50 | 132.95 | 1.55 | 221 | P03352 gallus gallus (chicke) |
| SwissProt_40:VE2_HPV1 + | 102.00 | 134.67 | 1.64 | 673 | P15116 homo sapiens (human) |
| SwissProt_40:VE2_HPV1 + | 102.00 | 132.41 | 1.69 | 509 | P05089 human papillomavirus t |
| SwissProt_40:SGS_DROE + | 101.50 | 138.65 | 1.66 | 660 | P03181 epstein-barr virus (str) |
| SwissProt_40:MANA_RHOM + | 101.50 | 137.91 | 1.95 | 297 | P000722 drosophila melanogast |
| SwissProt_40:VE2_HPV1 + | 101.00 | 133.17 | 1.97 | 1021 | P26445 rhodotermus marinus |
| SwissProt_40:ITLA_MOUSE + | 100.50 | 136.16 | 2.05 | 514 | P26445 human papillomavirus t |
| SwissProt_40:ITLA_MOUSE + | 100.50 | 129.98 | 2.23 | 336 | P11088 mus musculus (mouse) |
| SwissProt_40:ITLA_MOUSE + | 100.00 | 133.60 | 2.30 | 684 | P10600 aeromonas hydrophila |
| SwissProt_40:CCAB_RABIT + | 100.00 | 118.57 | 2.82 | 416 | P20930 homo sapiens (human) |
| SwissProt_40:EVAL_HUMAN + | 99.00 | 129.11 | 2.88 | 2339 | P05152 oryctolagus cuniculu |
| SwissProt_40:CCAB_HUMAN + | 99.00 | 117.97 | 3.38 | 592 | P09502 homo sapiens (human) |
| SwissProt_40:SFR_HUMAN + | 98.50 | 136.97 | 2.81 | 2231 | P000975 homo sapiens (human) |
| | | | | 221 | P001130 homo sapiens (human) |

| | | | | | |
|---------------------------|-------|--------|-------|------|------------------------------|
| SwissProt_40:VE2_HPV1 + | 97.00 | 128.59 | 4.02 | 454 | P080903 human papillomavirus |
| SwissProt_40:YB4_HSVH + | 97.00 | 127.54 | 4.07 | 512 | P28276 herpes simplex vir |
| SwissProt_40:YB4_SCHO + | 96.50 | 131.57 | 4.19 | 297 | P010357 schistosoma chromoc |
| SwissProt_40:CRE1_NEDOR + | 96.00 | 127.64 | 4.79 | 430 | P059958 neotropocera crassa |
| SwissProt_40:NONA_DROME + | 95.50 | 122.69 | 5.55 | 700 | P040047 drosophila melanog |
| SwissProt_40:VE2_HPV1 + | 95.00 | 125.79 | 5.77 | 452 | P36785 human papillomavirus |
| SwissProt_40:SFR_MOUSE + | 94.50 | 124.49 | 6.37 | 484 | P05519 homo sapiens (huma |
| SwissProt_40:KRI_CHICK + | 94.50 | 124.05 | 6.41 | 509 | P030373 gallus gallus (chl |
| SwissProt_40:UDB1_USUMA + | 94.50 | 122.75 | 6.52 | 591 | P07767 mus musculus (mous |
| SwissProt_40:DBA_HUMAN + | 94.00 | 118.62 | 6.89 | 950 | P40340 usliago maydis (s |
| SwissProt_40:GLT4_HUMAN + | 94.00 | 126.07 | 6.77 | 372 | P16988 homo sapiens (huma |
| SwissProt_40:YTA_RHIS + | 93.50 | 127.90 | 7.17 | 838 | P08489 trititum aestivum |
| SwissProt_40:SFR_MOUSE + | 93.00 | 127.03 | 7.86 | 278 | P17988 rhizolium sp. inser |
| SwissProt_40:GTO_HUMAN + | 93.00 | 119.82 | 8.66 | 283 | P060701 mus musculus (mou |
| SwissProt_40:GCE_HUMAN + | 93.00 | 118.17 | 8.85 | 784 | P10387 trititum aestivum |
| SwissProt_40:DBA_DROME + | 93.00 | 108.39 | 10.08 | 2411 | P12080 drosophila melanog |
| SwissProt_40:CCAA_HUMAN + | 93.00 | 113.15 | 9.46 | 2505 | P00055 homo sapiens (huma |
| SwissProt_40:YGSU_YEAST + | 92.50 | 129.43 | 8.26 | 198 | P40325 saccharomyces cere |
| SwissProt_40:YGR2_YEAST + | 92.50 | 122.56 | 9.05 | 436 | P00170 homo sapiens (huma |
| SwissProt_40:SFR4_HUMAN + | 92.50 | 121.68 | 9.18 | 494 | P08170 homo sapiens (huma |
| SwissProt_40:YK82_YEAST + | 92.50 | 113.98 | 10.15 | 1169 | P36170 saccharomyces cer |
| SwissProt_40:UW22_HUMAN + | 92.50 | 112.82 | 10.31 | 1336 | P015399 homo sapiens (hum |
| SwissProt_40:SGS3_DROE + | 92.00 | 111.93 | 11.32 | 1365 | P25172 drosophila melano |
| SwissProt_40:VE2_HPV1 + | 91.50 | 125.54 | 10.24 | 263 | P13728 drosophila yakuba |
| SwissProt_40:GATB_BOMO + | 91.50 | 120.06 | 11.01 | 494 | P36782 human papillomavirus |
| SwissProt_40:EBN1_EBV + | 91.50 | 119.82 | 11.05 | 508 | P32167 bombyx mori (silik |
| SwissProt_40:RL3_HAMA + | 91.00 | 125.63 | 11.09 | 641 | P03211 epstein-barr virus (|
| SwissProt_40:RLA_LENE + | 91.00 | 125.06 | 11.19 | 240 | P12706 halocaula marino |
| SwissProt_40:Y922_TREH + | 91.00 | 123.78 | 11.37 | 297 | P01200 leucocaula edodes (|
| SwissProt_40:VE2_HPV1 + | 91.00 | 119.28 | 12.07 | 498 | P08382 treponema pallidum |
| SwissProt_40:ATV4_HUMAN + | 91.00 | 107.99 | 13.72 | 1500 | P04656 human papillomavirus |
| SwissProt_40:Y236_HUMAN + | 91.00 | 109.96 | 14.04 | 1829 | P015015 homo sapiens (hum |
| SwissProt_40:U2R1_HUMAN + | 90.00 | 122.02 | 13.70 | 309 | P17143 human cytomegalovir |
| SwissProt_40:VE2_HPV1 + | 90.00 | 118.20 | 14.41 | 479 | P015695 homo sapiens (huma |
| SwissProt_40:VE2_HPV1 + | 89.50 | 118.04 | 14.45 | 488 | P26795 human papillomavirus |
| SwissProt_40:CMD4_HUMAN + | 89.50 | 113.69 | 16.61 | 742 | P16070 h cd4 antigen pre |
| SwissProt_40:POLN_RRVN + | 89.50 | 107.37 | 18.06 | 1533 | P12822 drosophila melanog |
| SwissProt_40:ZNA_RRVN + | 89.50 | 103.19 | 19.10 | 2479 | P13882 drosophila melanog |
| SwissProt_40:Y215_ADEO2 + | 89.50 | 96.46 | 20.89 | 5376 | P08799 mus musculus (mouse |
| SwissProt_40:Y215_ADEO2 + | 89.00 | 123.76 | 15.75 | 215 | P03691 human adenovirus t |
| SwissProt_40:Y215_ADEO2 + | 89.00 | 118.85 | 16.82 | 378 | P02833 drosophila melanog |
| SwissProt_40:Y215_ADEO2 + | 88.50 | 124.64 | 16.89 | 179 | P08893 herpes simplex vir |
| SwissProt_40:Y215_ADEO2 + | 88.50 | 123.81 | 17.03 | 197 | P15265 mus musculus (mouse |
| SwissProt_40:Y215_ADEO2 + | 88.50 | 113.36 | 19.63 | 655 | P01545 mus musculus (mouse |
| SwissProt_40:Y215_ADEO2 + | 88.50 | 110.77 | 20.32 | 882 | P01741 borrelia burgdorfer |
| SwissProt_40:Y215_ADEO2 + | 88.50 | 104.89 | 21.97 | 1733 | P33485 pseudorabies virus |
| SwissProt_40:Y215_ADEO2 + | 88.00 | 121.98 | 18.98 | 224 | P02872 gallus gallus (chi |
| SwissProt_40:Y215_ADEO2 + | 88.00 | 118.46 | 19.50 | 336 | P12574 xenopus laevis (chi |
| SwissProt_40:Y215_ADEO2 + | 88.00 | 112.58 | 21.51 | 660 | P04848 trititum aestivum |
| SwissProt_40:Y215_ADEO2 + | 88.00 | 109.29 | 21.68 | 704 | P17777 ratius norvegicus (r |
| SwissProt_40:Y215_ADEO2 + | 88.00 | 104.92 | 23.83 | 963 | P094572 ceenothabitis ele |
| SwissProt_40:Y215_ADEO2 + | 87.50 | 113.33 | 23.11 | 1593 | P28397 homo sapiens (hum |
| SwissProt_40:Y215_ADEO2 + | 87.50 | 109.04 | 24.47 | 558 | P000257 homo sapiens (huma |
| SwissProt_40:Y215_ADEO2 + | 87.50 | 108.33 | 24.70 | 914 | P38356 escherichia coli (e |
| SwissProt_40:Y215_ADEO2 + | 87.50 | 108.33 | 24.70 | 992 | P08563 rubella virus (str |

seq_name: SwissProt_40:HTRB_ECOLI
seq_documentation_block:
ID: HTRB_ECOLI STANDARD: PRT: 306 AA.
AC: P24187
DT: 01-MAR-1992 (Rel. 21, Created)
DT: 01-MAR-1992 (Rel. 21, Last sequence update)
DT: 16-OCT-2001 (Rel. 40, Last annotation update)
DE: Lipid A biosynthesis lauroyl acyltransferase (EC 2.3.1.-) (Heat shock
protein B).
GN: HTRB OR WAAW OR B1054 OR Z1690 OR ECG1432.
OS: Escherichia coli, and


```

398 TGGGGGGAACCTTACATGACGACGAGCTTCGTTCCATGACGCGCATG 447
|||||
138 euglyglylaryvalmetglyleucysglnpro.....MetMetAlaThr 152
|||||
448 TACAAAGCCGCGCAAAATGCAAGCATAGCAAAATTCATGACGAGGCGAC 497
|||||
153 tyrtargpronhisasnasnlnleumetglutrpvalglnthrargglylar 169
|||||
498 GGTGCGCGGCAAAAGCAAAACCGCGCCACCGGCATACAAAGGGGTCAAC 547
|||||
169 gmetarg.....SerAsnLysAlaMetLleGlyArgAsnMetLeuArgG 184
|||||
548 AATCATCATAGGCGCGCGCGCGGAGCAACCATCATCTCGCCGAC 597
|||||
184 LyllevalglalaleuylslysglgluAlaValAlaTrpPheAlaProsp 200
|||||
598 CACGTCCTCTCCGAGAGAGCGCGCGCTGTGGCGGATTTTTCGG 647
|||||
201 Gln.....AspTyrGlyArgLysSerSerPheAlaProPheAla 215
|||||
648 C.....AAACGTCATACACCATGACACTGCGGCAAAATTCGACACATCA 694
|||||
215 aValgluasnValAlaThrThrAsnGlyThrTyValleuSerArgLeuS 232
|||||
695 AAGCGGTGAAACCGTGTCTGCTGCGAACCGCTGCCGACGAGACA 744
|||||
232 erglyAlaAlaMetLeuThrValThrMetValArgLysAlaAspTyrSer 248
|||||
745 GCGTCGCTGTGGACATCGCGCGCGCTCAAGGGAATTAACGCG..... 789
|||||
249 GlyTyrArgLeuPheIleThrPro.....GlnMetGlnGlyTyrTr 262
|||||
790 .....ACAAAGCCGACGATGCGCGCGCTTCAACCGCAATACCG 829
|||||
262 oThrAspGlnAsnGlnAla.....AlaAlaTyrMetAsnLysLleIleG 277
|||||
830 AATATGTGATAGCGCGCTTCCGACGACGATGATGTTATGATCAACCGC 879
|||||
277 IulysGlnLleMetArgAlaProGlnLntyrLeuTrpIleHisArgArg 293
|||||
880 TATAAAGC 888
|||||
294 PheLysThr 296
|||||
seq_name: SwissProt_40:MSBB_HAEIN
seq_documentation_block:
ID MSBB_HAEIN STANDARD: PRT; 318 AA.
AC P44567;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lipid A biosynthesis (KDO)2-(lauroyl)-lipid IVA acyltransferase
DE (EC 2.3.1.-)
GN MSBB OR H10199.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OC NCBI_TaxID=727;
OK [1]
RN RP
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McInerney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shiley R., Liu L.-T., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;

```

```

RT "Whole-genome random sequencing and assembly of Haemophilus
RT Influenzae Rd."
RL Science 269:496-512(1995).
CC -1- FUNCTION: TRANSFERS MYRISTATE OR LAURATE, ACTIVATED ON ACP, TO
CC (KDO)2-(LAUROYL)-LIPID IVA (BY SIMILARITY).
CC -1- PATHWAY: LIPIDOLYSACCHARIDE CORE BIOSYNTHESIS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE HTRB/MSBB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isdb.ch/announce/
CC or send an email to license@isdb.ch).
CC -----
CC EMBL; U32705; AAC21868.1; -.
CC TIGR; H10199; -.
CC Lipopolysaccharide biosynthesis; Transferase; Acyltransferase;
CC Transmembrane; Inner membrane; Complete proteome.
KW TRANSMEM 27 47 POTENTIAL.
FT TRANSMEM 91 111 POTENTIAL.
FT TRANSMEM 138 158 POTENTIAL.
SQ SEQUENCE 318 AA; 36682 MW; DE59952D78719445 CRC64;

alignment_scores:
Quality: 168.00 Length: 265
Ratio: 1.105 Gaps: 10
Percent Similarity: 57.358 Percent Identity: 24.151

alignment_block:
US-09-303-518D-571 x MSBB_HAEIN ..
Align seg 1/1 to: MSBB_HAEIN from: 1 to: 318

112 CTGGGAAGACGCGCTCGACATCGCGCTTTCACCTTTAAGAGACCG 161
|||||
57 LendlylletprllegllyhislysAla.....LyslysglnAr 69
|||||
162 CGCGCGCATGCTGCGCAATATGCGGCGCG.....GGTTTGA 199
|||||
69 gThrArgAlaGlnThrAsnleuGlnTyrCysPhePronIstTrpGlnG 86
|||||
200 ACCCGACGACGACGCGCTGCAAGCCGTTTTCGGAACCGCAAAATGC 249
|||||
86 lnglnArgGlnGlnValLleAspLysMetPheAlaValAlaGlnVal 102
|||||
250 GCTTGGAACTTGCGCGCGCGCTTTCGCAAAACCGGAAACATCGAACC 299
|||||
103 MetPheGlyLleGlyGlnLleAlaLleArgSerLysLysHisLendlnly 119
|||||
300 AATGTTCAAGCGGTACAGCGCTGGGAACAGTGCGACGAGCTTGACA 349
|||||
119 sArgSerGlnPheLle.....GlyLeuGlnHisLleGlnGlnAlaLysAlaG 135
|||||
350 AGGGGGAAGGCGTGTGTCATGACGCGGACATCGGACATGAGATTGG 399
|||||
135 lueGlyLysAsnLleLleleuMetValPronHisGlyTrpAlaLleAspAla 151
|||||
400 GCGGAGCGCTACATC...AGCGACAGCTTCGTTCCACCTGACCGGCAT 446
|||||
152 serGlyLleLleLleuHisThrGlnGlyMetPro.....MetTrpSerMe 166
|||||
447 GTACAAAGCCGCGCAAAATCAAAAGCATGACAAATTCATGACGAGCGGCA 496
|||||
166 tyrtargpronhisasnasnlnleumetglutrpvalglnthrargglylar 183
|||||
497 GGTGCGCGGCAAAAGCAAAACCGCGCCACCGGCATACAAAGGGGTCAAA 546
|||||
183 rglnArgPheGlyGlyLysMetHisAlaArgGln...AsnGlyLleLys 198
|||||

```


| | | |
|--------------------------|-------------|-------------|
| DT | 01-MAR-1988 | |
| AC | P24205; | PRT; 323 AA |
| ID | MSBB_ECOLI | STANDARD; |
| 224-documentation_block: | | |

DT 01-MAR-1992 (Rel. 21, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Lipid 4 biosynthesis (Rel. 40, Last annotation update)
 DE (EC 2.3.1.1.)
 GN MSB or B1835.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
 NCBI_Taxid=562;

SEQUENCE FROM N.A.
STRAIN-K12 / W310;
MEDLINE-92121107; PubMed-1732206,
Kerov M., Georgopoulos C.;
"Isolation and characterization of the *Escherichia coli* msbA gene, a
multicopy suppressor of null mutations in the high-temperature
requirement gene *htrB*.";
J. Bacteriol. 174:702-710(1992).
[2]

SEQUENCE FROM N.A.
STRAIN-K12;
MEDLINE-93015688; PubMed-1355966;
Engel H., Smink A.J., van Wijngaarden L., Keck W.;
"Muclin-metabolizing enzymes from *Escherichia coli*:
second lytic transglycosylase";
J. Bacteriol. 174:6394-6403(1992).
[3]

SEQUENCE FROM N.A.
STRAIN-K12 / MG1655;
MEDLINE-97426617; PubMed-9276503.
Baltner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;
"The complete genome sequence of *Escherichia coli* K-12.";
Science 277:1453-1474(1997).
[4]

SEQUENCE FROM N.A.
STRAIN-K12:
MEDLINE=97251358; PubMed=9097040;
Itoh T., Aiba H., Bana T., Fujita K., Hayashi K., Inada T., Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K., Nakabe S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y., Sivasubramanian S., Tagami T., Takeda J., Takemoto K., Wada C., Yamamoto Y., Horichi T., A 460-kb DNA sequence of the *Escherichia coli* K-12 genome corresponding to the 40,150,000 min region on the linkage map. DNA Res. 3:379-392(1996).

RP FUNCTION, AND CHARACTERIZATION.
 RX MEDLINE=97256743; PubMed=9096727;
 RA Clementz T., Zhou Z., Raetz C.R.H.;
 RT "Function of the Escherichia coli msbB
 of htrB knockouts, in the acylation of
 lipid A. Acylation by MsbB

| | | |
|----|--|---|
| RT | | Follows laurate incorporation by HtrB.* |
| RL | J. Biol. Chem. 272:10353-10360(1997). | |
| CC | -1- FUNCTION: TRANSFERS MYRISTATE OR LAURATE, ACTIVATED ON ACP, TO THE | |
| CC | LIPID IVA MOIETY OF (KDO)2-(LAUROYL)-LIPID IVA. DECAOYL, | |
| CC | PALMITOYL, PALMITOLEOYL, AND (R)-3-HYDROXYMYRISTOYL-ACP ARE POOR | |
| CC | ACYL DONORS. FUNCTIONS OPTIMALLY AFTER LAURATE INCORPORATION BY | |
| CC | HTRB HAS TAKEN PLACE. ACYLATES (KDO)2-(LAUROYL)-LIPID IVA ABOUT | |
| CC | 100 TIMES FASTER THAN (KDO)2-LIPID IVA. DISPLAYS A PREFERENCE FOR | |
| CC | MYRISTOYL-ACP OVER LAUROYL-ACP. | |
| CC | -1- PATHWAY: LIPIDOLYSACCHARIDE CORE BIOSYNTHESIS. | |
| CC | -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane. | |
| CC | -1- SIMILARITY: BELONGS TO THE HTRB/MSB FAMILY. | |
| CC | -1- CAUTION: WAS ORIGINALLY (REF.2) THOUGHT TO BE THE MEMBRANE-BOUND | |
| CC | LYTIC MURIN TRANSGLYCOSYLASE (MLT). | |
| CC | ----- | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation. | |
| CC | the European Bioinformatics Institute. There are no restrictions on ways | |
| CC | used by non-profit institutions as long as its content is in no way | |
| CC | modified and this statement is not removed. Usage by and for commercial | |
| CC | entities requires a license agreement (see http://www.isb-sib.ch/announce/ | |
| CC | or send an email to license@isb-sib.ch). | |
| CC | ----- | |
| DR | EMBL; M77039; AAA24181.1; - | |
| DR | EMBL; M87660; AAA96706.1; - | |
| DR | EMBL; AE000279; AAC74925.1; - | |
| DR | EMBL; D90828; BAA15663.1; - | |
| DR | PIR; A42608; A42608. | |
| DR | CocGene; EG10614; msdb. | |
| KW | Lipopolysaccharide biosynthesis; Transferrase; Acyltransferase; | |
| FT | Transmembrane; Inner membrane; Complete proteome. | |
| FT | TRANSMEM 23 43 | POTENTIAL. |
| FT | TRANSMEM 85 105 | POTENTIAL. |
| FT | TRANSMEM 133 153 | POTENTIAL. |
| SO | SEQUENCE 323 AA; 37410 MW; 949DDA27770000 | |

| | | | |
|---------------------|--------|-------------------|--------|
| Quality: | 125.00 | Length: | 274 |
| Ratio: | 0.899 | Gaps: | 11 |
| Percent Similarity: | 50.730 | Percent Identity: | 24.088 |

alignment_block:

```

05 03-303-518D-571 x MSBB_ECOLI ..
Align seg 1/1 to: MSBB_ECOLI from: 1 to: 323

```

```

141 CCGCTCGGACATCTGGGCTTTTACCTTTAAAGGAAGACCGCGCGCAT 17
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
50 ArgLeuGIyArpPhealAGLyArgLeuGIySsrSerArGArGArgAl 66
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
171 CGTCGCCAATATG.....

```

```
66 aleuIIeasnlseuleCysPheproGluArgserCluaLacluar9g      :CGGC 187  
::: |||:::  
188 AGcGGGTtTGAAccCCAGCACgcAgAcGTCaaAAGcgTttTTggCGGA   ::|||  
:::|||                                                    |::|||  
83 luAla.....cttttttttttttttttttttttttttttttttttttttt    :|||||||
```

238 ACGCAAAATGCGGTTGSAACITGCCGCCGCGTTTCAAAAACCGGA 287
 :: :::::: ::::|
 93 AlaproginalametalametalagluLeuAlalleArgglyProgl 109

206 AGACATCAAGAACCAATGTTCAAAAGCGGTACACAGCGCTGGGACACGTCAGC 337
 |||||
 109 uLysIleGlnProArgValAsp...TripeIndlyLeuGlnIleIleGln 125
 |||||
 338 AGGCTTGACACAGGCGGAGGCGCTGCTGTTCAATCAGCGGCATATTCGCG 367
 ..

```

125  lumetAlgaTargasnasmGluPySvalIlePhenLeuValProHisGlyTyr 141
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
88  ACCTACGATTGGGCGGACCGCTACATCAGCAGCAGCGTTCCGTCCACCT 437
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

142 AlavalaIleProAlaMetLeuMetAlaSerGln...GlyInLysLe 157
 438 GACGGCCATGTCACAAAGCCGCAAAATCAAGGATAGCAAAATCATATC 487
 157 TAlaAlaMetPheHisSngInglYAsnProValPheAspTyrValTyrPA 174
 488 AGCGCGGAGGAGTGGCGGCAAGCAAAACCGCGCCACCGCATACAA 537
 174 snThyValArgArgArgPheGlyGlyArgLeuHisAlaArgAsn...Asp 189
 538 GGGGTCACAAATCATCAAGGCGCTGGCGCGCGGCGGCAACCATCAT 587
 190 GlyIleYAsnProPheIleGlnSerValArgInglYTrpGlyTyr 206
 588 CTGGCC...GACCACTCCCTTCCTCCGAGAGAGCGCGCGCTGT 631
 206 rLeuProAspGlnAspHisGlyProGlnHisSerGlu...P 219
 632 GGGGCGATTTTTCGCAAACTGATACACATGACACTGGCGGCAAA 681
 219 heValAspPhePheAlaThrTyrLysAlaThrLeuProAlaIleGlyArg 235
 682 TTGGCAGACGTC...AAAGCGTGAAGAACCTGTTTCTGCTGCGCA 725
 236 LeuMetLysValLysArgAlaArgValAlaProLeuPhe... 248
 726 AGCGCTGGCGGAGCAAA...GGCTTCGTGTTGACATCCGCGCGCG 769
 249 rProIleYAspGlyLysThrHisArgLeuThrIleGlnValAlaProP 265
 770 TCCAAAGGGATTTGACGCGCAAAAGCCGACAT...GCCGCGCTGTT 816
 265 rMetAspAspLeuLeuGlnAlaAspAspHisThrIleAlaArgArgMet 281
 817 AACCGCAATACCGAATATTGATACGCGCTTTCCGACGCGATTCGT 866
 282 AsnGlnGluValGlnIlePheValGlyProArgProGlnIleTyrThr 288
 867 TATGTACAAACCGTATTAACG 888
 298 rIleLeuLysLeuLeuLysThr 305
 seq_name: SwissProt_40:YGY3_HAISO

seq_documentation_block: PRT: 437 AA.
 ID YGY3_HAISO STANDARD: PRT: 437 AA.
 AC P21561,
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 50.6 kDa protein in the 5' region of GYRA and GYRB (ORF
 3).
 OS Haloferax sp. (strain Aa 2.2).
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloferax.
 OX NCBI_TaxID=2254;
 RN [1]
 RP MEDLINE=91100352; PubMed=1846146;
 RA Holmes M.L., Dyal-Smith M.L.;
 RT "Mutations in DNA gyrase result in novobiocin resistance in
 J. Bacteriol. 173:642-648(1991)."
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC EMBL: M38373; NOT_ANNOTATED_CDS.
 DR PIR: C39135; C39135.

KW Hypothetical protein.
 SQ SEQUENCE 437 AA; 50626 MW; B5B99A2AF3892BEF CRC64;
 alignment_scores: Length: 306
 Quality: 122.50
 Ratio: 0.996
 Percent Similarity: 40.196 Percent Identity: 23.529
 alignment_block:
 US-09-303-518D-571 x YGY3_HAISO ..
 Align seg 1/1 to: YGY3_HAISO from: 1 to: 437
 27 TCCGCGCTTCGCAAGCGGATCCATCCCTGTGACCGCCCTGCTCAAT 76
 130 SerArgLeuArgGlnGlnHisGlnHisProArgGlyArg..... 142
 77 GCGTCCCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 149
 143HisAlaSerAspArgValG 149
 127 GACATCTGGCGGTTTACCTTTAAAGAGAGCGCGCGCATGCTGCG 176
 149 In...AspGlyAlaHisProArgArgGlnArgLeuArgGlu..... 161
 177 CATATGCGCGAGCGGCTTGAACCGCGACGCGACGAGCGTCAAGCGG 226
 162GlnProArgHisAla..... 166
 227 TTTTTCGGAAGCGCAAAATGCGGTTTGAACCTTGGCCCGCGGTTTTC 276
 167GlyArgProArgArgArgGlnProProArgArgGlyArg 179
 277 AAAAACCAGAAACATCAATGTTCAAGCGGATACGCGCTGGA 326
 179 rSerArgGlyThrHisArgArgHisLeuArgGlnAlaProArgProAla 195
 327 ACACGTCGACGAGCTTGGACAAAGCGGAGCGCTGCTGCTCATCAGC 376
 196 ValArgGlyProAspGlnAspGlnAlaArgGlu..... 206
 377 GCGACATCGGAGCTACGATTTGGCGAGCGCTACATCAGCGAGCTT 426
 207PheArgGlyProArgHisArgArgGlnArgHisProProAla. 221
 427 CGGTCCACCTGACCGCATGTACAAAGCGCGCAAAATCAAGCGATAGA 476
 222ArgAspValLeuArgGlyGlnProGlnHisGlyAsp 233
 477 CAATATCATGAGCGCGGAGGATGCGCGCAAAAGCGCGCGCA 526
 234 GlyHisHisLeuGlnGlyArg..... 240
 527 CCGGATACAAAGGCTCAACAAATCATCAAGGCGCGCGGAG 576
 241ArgGlyArgProArgProGlnGlyArgGlnAlaGlyArg 254
 577 GCAACATCATCT... 590
 254 LysAlaHisProProGlnValArgAlaArgIleTyrLeuAlaAlaGlyGlu 270
 591GCCGACCGAGTCCCTTC..... 608
 271 AlaArgGlyLeuProGlnProArgProLeuGlyValArgThrValHisAr 287
 609TCGCGAGAGAGCGCGCGCGCTG 630
 287 gGlyGlyArgLeuArgGlyArgValGlnAlaGlyProArgProGlnV 304
 631 TGGCGGATTTTTCG...CAAAACGTCATACCATGACACT 671
 304 alProGlyAspPheAlaProGlnGlyGlnAspSerGlnArgArgGlnThr 320


```

672 GCGGCAAAATTCGACACGTCAAGGCGTGAACCCCTGTTTCTGCT 721
    ::::::::::::::::::::::::::::
321 ProProAlaProHisSerArgLysArgArgAspThrValAlaHisHisAr 337
722 GCGAAGCGCTCCCGACGAGCAAGCGCTGCT..... 752
    ::::::::::::::::::::::::::::
337 GHisTrpArgArgArgArgArgValAlaArgHisArgGluGlyAlaLeuP 354
753 ..GTTGCACATCCGCGCCCTCCAAAGGGAATTGAAGCGCAACCCCA 800
    ::::::::::::::::::::::::::::
354 roAlaAlaHisProAspArgArgArgArgArgArgArgAlaHisPro 370
801 CGATGCGCGCGCTGCTTC 816
    ::::::::::::::::::::::::::::
371 ASPAlaAlaAlaLys 375
seq_name: SwissProt_40:MSBB_SHIFL

```

```

seq_documentation_block:
ID MSBB_SHIFL STANDARD: PRT; 231 AA.
AC 006559:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Lipid A biosynthesis (KDO)2-(lauroyl)-lipid IVA acyltransferase
  (EC 2.3.1.-) (Fragment).
GN MSBB.
OS Shigella flexneri.
OC Plasmid 230 kb pM5H6000.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
  Shigella.
OX NCBI_TaxID=623;
  [1]
RN SEQUENCE FROM N.A.
RC STRAIN=2A;
RX MEDLINE=97315241; PubMed=9171415;
RA Rededge L., Davis M.A., Youngren B., Austin S.J.;
RT "Plasmid maintenance functions of the large virulence plasmid of
  Shigella flexneri."
RL J. Bacteriol. 179:3670-3675(1997).
CC -|- FUNCTION: TRANSFERS MYRISTATE OR LAURATE, ACTIVATED ON ACP, TO
  (KDO)2-(LAUROYL)-LIPID IVA (BY SIMILARITY).
CC -|- PATHWAY: LIPIDOLYSACCHARIDE CORE BIOSYNTHESIS (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
  (Potential).
CC -|- SIMILARITY: BELONGS TO THE HTRB/MSBB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U82621; AAB58154.1;
KW Lipopolysaccharide biosynthesis; Transferase; Acyltransferase;
  Transmembrane; Inner membrane; Plasmid.
FT NON_TER 1
FT TRANSMEM 48 68 POTENTIAL.
SQ SEQUENCE 231 AA; 26754 MW; 99350DB0A499D854 CRC64;

```

```

alignment_scores:
  Quality: 120.50      Length: 249
  Ratio: 0.996        Gaps: 10
  Percent Similarity: 48.594      Percent Identity: 22.088

```

alignment_block:

us-09-303-518d-571 x MSBB_SHIFL

Align seg 1/1 to: MSBB_SHIFL from: 1 to: 231

```

217 GTCAAGCCGTTTTTGGGAAAGCGCAAAATGCGGTTTGGAAGTTGCCCC 266
    ::::::::::::::::::::::::::::
1 ValAspAsnMetPheAlaThrAlaLeuGlnSerIleValaMetMetLeu1 17
267 GCGCTTTTCAAAAACCGCAAGACATCGAACATATGTTCAAGCGGTA. 315
    ::::::::::::::::::::::::::::
17 uLeuAlaIleArgGlyProGluLysPheGln.....LysArgValaP 31
316 ....CAGCGCTGGAAACAGCTGCACAGCGCTTGGACAAAGCGGCAAGG 360
    ::::::::::::::::::::::::::::
31 heTrpLysGlyLeuGluIleLeuGluGluIleArgHisAsnAsnArgAsn 47
361 CTGCTGTTCATCAGCGCGGCAATCGGACAGCTAGCATTTGGCGGCGGCTA 410
    ::::::::::::::::::::::::::::
48 ValIlePheLeuValProHisGlyTrpSerValaAspIleProAlaMetLe 64
411 CATCAGCCAGCAAGCTTCCTGCACCTGACCGCGCATGTACAAAGCGCGCA 460
    ::::::::::::::::::::::::::::
64 uLeuAlaAlaGln...GlyGluLysMetAlaAlaMetPheHisGlnGlnA 80
461 AATCAAGCGATAGCAAAATCATGCAAGCGGCGAGGTCGCGCGCAAA 510
    ::::::::::::::::::::::::::::
80 rGAsnProValIleAspTrpValaIleProAsnSerValaArg...ArgLysPhe 95
511 GCGAAACCGCGCGCCAGCGCATACAAAGGCTCAACAAATCATCAAGCG 560
    ::::::::::::::::::::::::::::
96 GlyGlyArgLeuGlnHisSerArgGluAspGlyLysProPheIleGlnSe 112
561 CTGCGCGCGGCGGCGAGCAATCATCTGCTCC.....GACCAGCTCC 604
    ::::::::::::::::::::::::::::
112 rValArgGlnGlyTrpTrpGlyTrpLeuProAspGlnAspHisGlyP 129
605 CTTCGCGCAGAAAGCGCGCGCTGCGCGGATTTTTCGCGCAAGCT 654
    ::::::::::::::::::::::::::::
129 rGluTrpLysSerGlu.....PheAlaAspPheAlaThrTrp 141
655 GCATACACACATGACAGCGCGGCAAAATTTGGCACAGCTCAAGCGCTGA 704
    ::::::::::::::::::::::::::::
142 LysAlaThrLeuProIleIleGlyArgLeuMetAsnIle..... 154
705 AACCTGTTTTTGTCTGCACAGCGCTGCGCGAGCAAGCGCTTCGTCT 754
    ::::::::::::::::::::::::::::
155 .....SerGlnAlaMetIleI 160
755 TGCACATCGCGCGCTGCAGGGAATTGAAGCGCAACCAAGCCGAC... 801
    ::::::::::::::::::::::::::::
160 leProLeuPheProValaIleAspGlu.....LysLysHisPhe 172
802 .....GATGCGCGCGCTG..... 813
173 LeuThrIleGluValaArgProPheMetAspAlaCysIleAlaSerAlaAs 189
814 .....TTCACCGCGAATCCGCAATATGGATAC 841
    ::::::::::::::::::::::::::::
189 rAsnLysMetIleAlaArgGlnMetAsnLysThrValaGluLeuValaG 206
842 GCGGTTTCCGACGAGTATGTTATGTATCAACCGCTATAAAGC 888
    ::::::::::::::::::::::::::::
206 LysHisProGluGlnTrpIleTrpValaLeuLysLeuLysThr 221
seq_name: SwissProt_40:NFX1_HUMAN

```

```

seq_documentation_block:
ID NFX1_HUMAN STANDARD: PRT; 1104 AA.
AC Q12986:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Transcriptional repressor NF-X1 (Nuclear transcription factor, X box-
  binding, 1).
GN NFX1.
OS Homo sapiens (Human).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95053707; PubMed=7964459;
 RA Song Z., Krishna S., Thanos D., Strominger J.L., Ono S.J.;
 RT "A novel cysteine-rich sequence-specific DNA-binding protein
 RT interacts with the conserved X-box motif of the human major
 RT histocompatibility complex class II genes via a repeated Cys-His
 RT domain and functions as a transcriptional repressor."
 RL J. Exp. Med. 186:1763-1774(1994).
 CC -1- FUNCTION: REPRESSOR OF HLA-DRA TRANSCRIPTION. BINDS TO THE X-BOX
 CC MOTIF OF CLASS II MHC GENES. MAY PLAY AN IMPORTANT ROLE IN
 CC REGULATING THE DURATION OF AN INFLAMMATORY RESPONSE BY LIMITING
 CC THE PERIOD IN WHICH CLASS II MHC MOLECULES ARE INDUCED BY
 CC INTERFERON-GAMMA.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- INDUCTION: BY INTERFERON GAMMA.
 CC -1- SIMILARITY: TO D.MELANOGASTER SHUTTLE CRAFT PROTEIN (STC) AND
 CC YEAST YNL023C.
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U15306; AAA69517.1;
 DR HSSP: P02876; ZMGC.
 DR MIM: 603255;
 DR InterPro: IPR001965; PHD.
 DR InterPro: IPR001374; R3H.
 DR InterPro: IPR000967; ZNF_NFX1.
 DR InterPro: IPR001841; Znf_ring.
 DR Pfam: PF01424; R3H; 1.
 DR Pfam: PF01422; zf-NFX1; 8.
 DR SMART: SM00393; R3H; 1.
 DR SMART: SM00184; RING; 1.
 DR SMART: SM00438; ZNF_NFX; 9.
 DR PROSITE: PS00518; ZF_RING_1; FALSE_NEG.
 DR PROSITE: PS0089; ZF_RING_2; 1.
 DR Transcription regulation, Repressor; DNA-binding; Nuclear protein;
 KW Repeat; Zinc-finger.
 KW Repeat; Zinc-finger.
 FT ZN_FING 342 393 RING-TYPE.
 FT DOMAIN 424 876 7 X APPROXIMATE REPEATS, CYS-RICH.
 FT REPEAT 424 463 1.
 FT REPEAT 480 517 2.
 FT REPEAT 541 580 3.
 FT REPEAT 606 647 4.
 FT REPEAT 695 734 5.
 FT REPEAT 806 842 6.
 FT REPEAT 843 876 7.
 SO SEQUENCE 1104 AA; 123149 MW; 36FEC4292F78130C CRC64;

 alignment_scores: Length: 374
 Quality: 119.00 Gaps: 17
 Ratio: 0.937 Percent Identity: 19.786

 Percent Similarity: 33.957 Percent Identity: 19.786

 alignment block:
 US-09-303-518D-571 x NFX1_HUMAN ..

 Align seg 1/1 to: NFX1_HUMAN from: 1 to: 1104

 14 AATTGAGCTGTTTCCCTTTCGACGAGG...CCATGCACATCCGCT... 58
 ||||| ::::::::::: ||||| ::::|
 553 AsnHisThrCysSerGlnValCysHisProGlnProCysGlnGlnCysPrr 569

59TGACGGCCCTGCTCAATATCCCTCCCTGCTGCTTCCCTGTCT 104
 ||||| ::::::::::: ||||| ::::|
 569 oATgLeuProGlnLeuValArgCysCysProCysGlnGlnThrProLeuS 586
 105 GCACACGCTGGGAAACCGCTCGGACATCTGGCGTTTACCTTTAAAGG 154
 ||||| ::::::::::: ||||| ::::|
 586 ergInLeuLeuGlnLeuGlnSerSerSer.....Arg 596
 155 AAGACCGCGCGCGCATCTGTCGCAATATGCGCGCGGTTTGAACCC 204
 ||||| ::::::::::: ||||| ::::|
 597 LysThrCysMetAspProValProSerCysGly..... 607
 205 GACACGCGACGCTCAAGCGCTTTTTCGGAACGCGCAAAATGCGGTT 254
 607
 255 GGAACCTGCCCCCGCTTTTCAAAAACGGAAACATGAAACATGT 304
 ||||| ::::::::::: ||||| ::::|
 608LysValCys 610
 305 TCAAGCGGTACACGCGTGGG.....AACGCTGCGACG 339
 ||||| ::::::::::: ||||| ::::|
 611 GlyLysProLeuProCysGlySerLeuAspPheIleHisThrCysGly 627
 340 GCTTTGGACAGGCGGAAAGGCTGCTTCATCAAGCGCGCATCGGCGAG 389
 ||||| ::::::::::: ||||| ::::|
 627 sleu...CysHisGlnGlnAspCysGlyProValSerArgThrVal. 642
 390 CTACGATTTGGCGGACGCTACATCAGCAGCAGCTTCCTCCACTGA 439
 ||||| ::::::::::: ||||| ::::|
 643IleSerCysArgCysSerPheArgThrLysGlu 653
 440 CGCGCATGTACAAAGCCCGCAAA.....TCAAA 468
 ||||| ::::::::::: ||||| ::::|
 654 LeuProCysThrSerLeuLysSerGlnAspAlaIleThrPheMetCysAsp 670
 469 GCGATGACAAAA.....TCATGCAAGCGGCGAG 497
 ||||| ::::::::::: ||||| ::::|
 670 sATgCysAsnLysLysArgLeuCysGlnArgHisLysCysAsnGluIleC 687
 498 GGTGCGCGGCAAGGCAAAACCGCGCCA..... 526
 ||||| ::::::::::: ||||| ::::|
 687 yscysValAspLysGlnHisLysCysProLeuAsnCysGlnArgLysleu 703
 527CGGCATACAGGCGTCAACAA 549
 ||||| ::::::::::: ||||| ::::|
 704 ArgCysGlnLeuHisArgCysGlnGluProCysHisArgGlnLysCysG 720
 550 ATCA..... 553
 ||||| ::::::::::: ||||| ::::|
 720 nThrCysTrpGlnAlaSerPheAspGlnLeuThrCysHisCysGlnAla 737
 554TCAGGCGCC..... 562
 ||||| ::::::::::: ||||| ::::|
 737 ervaIleLeuTrpProValProCysGlnThrArgProGluCysThr 753
 563TCGCGCGCGGCGAGGCA..... 580
 ||||| ::::::::::: ||||| ::::|
 754 GlnThrCysAlaArgValHisGlnCysAspHisProValTyHisSerG 770
 581CCATCATCTGCGCGACACGCTGCTTCGCGAGAGAGCGG 623
 ||||| ::::::::::: ||||| ::::|
 770 yHisSerGlnGlnLysCysProCysThrPheLeuThrGlnLys.... 785
 624 CGCGGTGTGGCGGATTTTTCGGCAACCGCANACACATGACACTGG 673
 ||||| ::::::::::: ||||| ::::|
 786Trr 786
 674 CGGCAAAATTTGGACACAGCTCAAGAGCGTGAACAACTGTGTTT..... 715
 ||||| ::::::::::: ||||| ::::|
 787 CysMetCysLysHisGlnLysLeuArgSerAsnIleProCysHisLeuValAs 803
 716TCGTCTGCGAAGCGCTGCGCG..... 736

```

803 p15serCysGlyLeuProCysSerAlaThrLeuProCysGlyMetHisL 820
737 .....ACGACAAAGCCTTCGCTTGCACATCCGCCCTGCC 772
820 yscysGlnAlaGlyLeuCysHisLysGlyLeuValAlaSpGlnProCys 836.
773 AAGGGGAATTGACGGCAACAAG.....CCGACGATGCC 807
837 LysGlnProCysThrThrProArgAlaAspCysGlyHisProCysMetAl 853
808 GCCCTGTTCACACCGCATACCG 829
853 aProCysHisThrSerSerPro 860

seq_name: SwissProt_40:HKR3_HUMAN

seq_documentation_block:
ID HKR3_HUMAN STANDARD; PRT; 688 AA.
AC P10074;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Kruppel-related zinc finger protein 3 (HKR3 protein).
GN HKR3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RA MEDLINE=95059073; PubMed=7969177;
R Sugawara M., Scholl T., Ponath P.D., Strominger J.L.;
RT "A factor that regulates the class II major histocompatibility
RT complex gene DPA is a member of a subfamily of zinc finger proteins
RT that includes a Drosophila developmental control protein.";
RL Mol. Cell. Biol. 14:8438-8450(1994).
[2]
RN
RP SEQUENCE FROM N.A.
RA MEDLINE=98177483; PubMed=9516840;
R Matis J.M., Jensen J., Sulman E.P., Beltinger C.P., Allen C.,
R Biagel J.A., Brodeur G.M., White P.S.;
RT "Human Kruppel-related 3 (HKR3): a candidate for the 1p36
RT neuroblastoma tumour suppressor gene?";
RL Eur. J. Cancer 33:1991-1996(1997).
[3]
RN
RP SEQUENCE OF 461-488 FROM N.A.
RA MEDLINE=89096896; PubMed=2850480;
R Ruppert J.M., Kinzler K.W., Wong A.J., Bigner S.H., Kao F.T.,
R Law M.L., Seunasz H.N., O'Brien S.J., Vogelstein B.;
RT "The G1I-Kruppel family of human genes.";
RL Mol. Cell. Biol. 8:3104-3113(1988).
[4]
RN
RP -1- FUNCTION: BINDS TO AND REGULATES THE J AND/OR S ELEMENTS IN MHC II
RP PROMOTER.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L16896; AAA65124.1; -
CC EMBL; U45325; AAB08973.1; -
CC EMBL; U45324; AAB08973.1; JOINED.
CC EMBL; M20677; AAA35989.1; -
CC PIR; E31201; E31201.
CC HSSP; P08047; ISP2.

```

```

DR MIM; 165270; -
DR InterPro; IPR000210; BTB_POZ.
DR InterPro; IPR000822; ZnF-C2H2.
DR Pfam; PF00651; BTB; 1.
DR Pfam; PF00096; ZF-C2H2; 11.
DR PRINTS; PR00048; ZINCFINGER.
DR SMART; SM00225; BTB; 1.
DR SMART; SM00355; ZnF-C2H2; 11.
DR PROSITE; PS50097; BTB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 9.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 11.
DR Transcription regulation; Activator; Zinc-finger; DNA-binding;
KW Repeat; Metal-binding; Nuclear protein.
FT DOMAIN 26 89
FT ZINC FINGERS.
FT ZN_FING 291 600
FT ZN_FING 291 313
FT ZN_FING 319 342
FT ZN_FING 350 372
FT ZN_FING 378 401
FT ZN_FING 407 430
FT ZN_FING 436 459
FT ZN_FING 465 487
FT ZN_FING 493 515
FT ZN_FING 521 544
FT ZN_FING 550 572
FT ZN_FING 578 600
FT CONFLICT 201 201
FT CONFLICT 244 244
FT CONFLICT 350 351
FT CONFLICT 607 607
SQ SEQUENCE 688 AA; 77054 MW; EBECCE3D6CB524 CRC64;

```

```

alignment_scores:
  Quality: 117.50      Length: 241
  Ratio: 1.078        Gaps: 14
Percent Similarity: 45.228      Percent Identity: 25.311

```

alignment_block:

US-09-303-518D-571 x HKR3_HUMAN ..

Align seg 1/1 to: HKR3_HUMAN from: 1 to: 688

```

84 CCGCTGCTGCGCTTCGCTGCGACAGCGTGGAAACCGCT.....125
||||| .....
133 ProAlaSerGlnAsnValAlaSerHisValLysIleProAlaGlyLeuG1 149
126 .....CGACATCTGCGCTTACCTTTAAAGAGAGAC 159
149 uGlUGlUGlValAlaSerArgThrLeuGlyLeuValProArgAspGlnIup 166
160 CGCGCGGCATCGTCGCCCATATCCGCGAGCGGCTTGAACCCGACAC 209
||||| .....
166 roAlaGlySerHisSerProGlnArgProGlnLeuHisSerProAlaGln 182
210 GCACAGCGTCAAAAGCCGTTTTCGGAAGCGCAAAATGCGGTTTGAAC 259
||||| .....
183 SerLysGlyProSerSerLeuGlyGlyLeuLysGlnAlaLeuLyspr 199
260 TTGCCCCCGGCTTTTCAAAAACCGGAGAGC.....291
||||| .....
199 oCys.ProlaGluAspLysLysProGlnLysPysLysValProporArg 215
292 ...ATCGAAACATGTTCAAGCGGTACAGGC.....TGGA 326
||||| .....
216 ProlaGluAlaGluGlyAlaGlnLeuGlnGlySerAsnGluIupr 232
327 ACACGTGCACAGCTTTGCGACAGGCGAAGGCGTGTTCATCAGCG 376
||||| .....
232 uValValAlaGlnValGlnLysPaspGlyasp.....242
377 CGCACATCGCAGCTACGATTTGGCGGAGCGTACATCAGCCAGAGCTT 426
||||| .....

```

```

243 ..... GlyaspTyrMetSerGluPro... 249
427 CCGTTCACCTGACCGCATGTACAGCCCGCAAAATCAAGCATAGA 476
250 ..... GluAlaVal.LeuThrArgGlySer..... 258
477 CAAATCATGACGAGCGGAGGTGGCGGCAAGGCAAAACCGGCCCA 526
259 ..... AsnValIleArgLysProCysAlaIleGlu..... 268
527 CCGCATCATCAGGGGTCAAAATCATCAAGGCGTGGCGGCGGCGAG 576
269 ProAlaLeuSerIleArgLysLeuAlaIleGluProAlaGluAsnArgLys 285
577 G..... CACCATCATCTGCGCCGACACGATCCCTTCTCCGACAGA 617
285 sglThrAlaValProValGluCysProThr..... 295
618 AGCGGCGGCGCTGTGGGCGGATTTTTCGCAACGTCATACACATGA 667
296 ..... CysHisLysLysPheLeuSerLysTyrTyrLeuLysVal 308
668 CACTGCGCGCAAAATTTGGCAGACGTCAAGGCGTGAAGAACCTGTTTC 717
309 HisAsnArgLys..... HisThrGlyLys..... LysProPhe..... 319
718 TGCTGCGAAGCGCTGCCCG 736
320 ..... GluCysPro 322

seq_name: SwissProt_40:DBPA_RAT

seq_documentation_block:
ID DBPA_RAT STANDARD; PRT; 361 AA.
AC Q62764; Q63748;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA-binding protein A (Cold shock domain protein A) (Muscle Y-box
  protein YB2) (Y-box binding protein-A) (RYB-A).
GN CSDA OR DBPA OR YB2
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP STRAIN=SPRAGUE-DAWLEY;
RA Goldaman D., Gao J., Burmeister M., Sapru M.;
RT "Characterization of muscle Y-box proteins that bind the mACHR
  delta subunit promoter."
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=MISTAR; TISSUE=Liver;
RX MEDLINE=94301785; PubMed=8029009;
RA Ito K., Tsutsumi K., Kuzumaki T., Gomez P.F., Otsu K., Ishikawa K.;
RT "A novel growth-inducible gene that encodes a protein with a
  conserved cold-shock domain."
RL Nucleic Acids Res. 22:2036-2041(1994).
CC -1 SUBCELLULAR LOCATION: Nuclear.
CC -1 ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS: 1 (SHOWN HERE) AND 2;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1 TISSUE SPECIFICITY: ABUNDANT IN THE SKELETAL MUSCLE, SPLEEN, AND
  FETAL LIVER.
CC -1 SIMILARITY: BELONGS TO THE COLD-SHOCK DOMAIN (CSD) FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (see http://www.isb.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC EMBL; U22893; AAB60520.1; -.
DR EMBL; D28557; BAA05907.1; -.
DR HSSP; P15277; IMJC.
DR InterPro; IPR002059; Cold_shock.
DR Pfam; PF00313; CSD; 1.
DR PRINTS; PR00050; COLDSHOCK.
DR PRODOM; PD00621; Cold_shock; 1.
DR SMART; SM00357; GSP; 1.
DR PROSITE; PS00352; COLD_SHOCK; 1.
KW Transcription regulation; Repressor; DNA-binding; Nuclear protein;
  Alternative splicing.
FT DOMAIN 85 149 CSD.
FT VARSPITC 184 252 MISSING (IN ISOFORM 2).
FT CONFLICT 14 14 L->H (IN REF. 2).
FT CONFLICT 52 74 SPGGDADGAPAPASSAPGSEDA -> APARASPRARPGIL
  SPRKRRR (IN REF. 2).
FT PR -> HV (IN REF. 2).
FT CONFLICT 120 121 PR -> HV (IN REF. 2).
FT SEQUENCE 361 AA; 38851 MW; C6799D5A3DA5C3F3 CRC64;
SQ

```

```

alignment_scores:
  Quality: 110.50      Length: 201
  Ratio: 1.315        Gaps: 8
  Percent Similarity: 41.791      Percent Identity: 24.876

```

```

alignment_block:
US-09-303-518D-571 x DBPA_RAT ..

```

```

Align seg 1/1 to: DBPA_RAT from: 1 to: 361

```

```

147 TTAAAGAGAGACCGCGCGCATGCTGCCAATATGCG..... 185
173 TyrTyrGlyArgArgArgGlyProProArgAsnTyrAlaGlyGluGlu 189
186 ..... GCAGCGGTTTGAACCCGACAGCAGCAGAG 216
189 uGluGluGlySerGlySerSerGluGluGlyPheGluProProAlaAlaSpG 206
217 GTCNAAGCCGTTTTCGGAACGCGCAAAATGCGG..... 251
206 LysIlePheSerGlyAlaArgAsnGlnLeuArgArgProGlnTyrArgPro 222
252 ..... TTGG 256
223 ProTyrArgLysArgPheProProTyrHisValGlyIleThrPheAs 239
257 AACTTGCCCGCGCGTTTTCAAAAACCGAAGACATCGAACAATGTT 306
239 PArgArgSerArgValPhePro..... HisProAsnArgMetG 252
307 AAACCGTACAGCGCTGGGAACAGCTGACAGCGCTTTGGACAAGGCGA 356
252 IlnIleGly... GluIleGlyIleMetLysAspGlyValProGluGlyAla 267
357 AGGCGTCTGTTTCATCAAGCGCGACATCGCAGCTACAGATTGGCGGAC 406
268 GlnLeuGlnValHisArgAsnProThrTyrArgProArgPheArgGly 284
407 GCTACATCAGCCAGCAGCTTCGTTCCACCTGACCGCATGTACAACCG 456
284 YProAlaArgPro..... ArgProAlaProAlaI 294
457 CCGAAATCAAGGATGACAAATCATGACGCGGCGGAGGTGCGCGG 506
294 IeGlyIleAlaGluAspLysGluAsnGlnGlnAlaIleAsnGlyProAsn 310
507 CAAGAGCAAAACCGCGCCACCGGCATACAGAGGGTCAACAAATCATCA 556
311 Gln..... 311
557 AGGCCCTGCGCGGCGGCGAGGCAACATCATCT..... GCCC 594

```

```

312 .....ProSerAlaArgGlyPheArgArgProTyrAsnTyrArgArgA 327
595 GACCAGTCCTCCCTCTCCGCA.....GGAAGCGCGCGCGTGTG 632
327 rgrProArgProLeuAsnAlaValSerGlnAspGlyLysGlnThrLysAla 343
633 GGC 635
344 Gly 344

```

seq_name: SwissProt_40:VE2_HPV19

seq_documentation_block:

```

ID VE2_HPV19 STANDARD; PRT; 493 AA.
AC P36786;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Regulatory protein E2.
E2
OS Human papillomavirus type 19.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10608;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=94265501; PubMed=8205838;
RA Dollus H., Hofmann B.;
RT "Primer-directed sequencing of human papillomavirus types.";
RT Curr. Top. Microbiol. Immunol. 186:13-31(1994)
CC -1- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACNNNNNGGT-3') PRESENT
CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING ON E2RE'S POSITION
CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC BY SPECIALLY HANDLING THE ASSEMBLY OF THE TRANSCRIPTION
CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC REPLICATION.
CC -1- SUBUNIT: BINDS DNA AS A DIMER.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X74470; CAAS3521.1; -
DR PIR: S36488; S36488.
DR HSSP: P03122; ZBOP.
DR InterPro: IPR001866; E2_N.
DR InterPro: IPR000427; Early2_C.
DR Pfam: PF00511; E2_C; 1.
DR Pfam: PF00508; E2_N; 1.
DR ProDom: PD000672; Early2_C; 1.
DR ProDom: PD000678; E2_N; 1.
DR Early protein; transcription regulation; Activator; DNA-binding;
KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
SQ SEQUENCE 493 AA; 55740 MW; DF8F316FC66C6782 CRC64;

```

alignment_scores:

```

Quality: 110.50 Length: 244
Ratio: 0.929 Gaps: 8
Percent Similarity: 48.770 Percent Identity: 23.361

```

alignment_block:

```

US-09-303-518d-571 x VE2_HPV19
Align seg 1/1 to: VE2_HPV19 from: 1 to: 493

```

```

200 ACCCCGACACGACGACGCTCAAGCCGTTTTTTCGGAACGGCAAAATGC 249
224 ThrAspSerAlaSerArgLeuSerProThrAlaSerArgLysGlnSerG1 240
250 GGTTCGTAACCTTGCCCGCCGCTTTTCAAAAACCGGAACGACATGAAA 298
240 ngInThrAsnThrLysGlyArgArgTyrGlyArgArgProSerSerArgT 257
299 .....CATGTTCAAGCGGTACAGCGGCGGGAACACGTCG 334
257 hrArgArgGlnThrGlnThrArgGlnLysArgSerArgSerLysSerLys 273
335 ACCAGCGTTTGGACAAGGCGGAGGCGCTGCTTCATCAGCGCGACATC 384
274 SerArgSerArgSerArgSerArgSerArgSerArgSerArgSerArgSer 290
385 GGCAGCTTCAGATTGGCGGAGCGCTACATCAGCCAGCAGCTTCGCTCCA 434
290 gSerLysSerArgArgLysAlaSerThrThrArgGlyArgGlyArgGlys 307
435 CCGTACCGCCATGTACAAAGCCGCGAATCAAGCGATGACAAATCA 484
307 er...ProThrAlaThrSerArgSerArgSerArgSerArgSerProSerAlaThr 322
485 TGCAGCGCGGCGGCGGCGCGCAAGCAAAACCGCGCCACCGGCATTA 534
323 .SerSerThrThrSerLeuArgSerArgLysSerArgValGlyArgS 339
535 CAAGGGGCAACAAATCATCAAGCGCGCGCGCGGCGGAGCAACCAT 584
339 erArgGlyGlyArgSerArgValGlyArgSerArgLysArgLysArg 355
585 CATCTGCCCGACACGCTCCCTCCGAGAGAGCGGCGGCGTGTGG 634
356 SerArgLysSerProSerProThrAsnThrLysArgSerArgArgLys 372
635 CGGATTTTTCG. ....CAACCTGCA 657
372 rGlySerSerArgLeuHisGlyValSerAlaAspAlaValGlyThrSerV 389
658 TACACCATGACATGCG.....GGCAAAATGCG.....ACAGCT 692
389 alHisThrValSerGlyArgAsnThrGlyArgLeuGlyArgLeuLeu 405
693 CAAGGCGTGAACCCGCTTTTTCGCTGCA.....ACGCTGC 733
406 GluAlaLeuAspProValIleLeuValArgGlyLysProAsnThrIle 422
734 CCGACGACAAAGCGTTCGTTGCACATCGCCCGCTGCAAGGGGAATTG 783
422 uArgSerPheArgAsnArgAlaLysHis..... 431
784 AACGGCAACAAAGCCACGATCGCGCGTTCACCCGCAATACCGAATA 833
432 .....MetTyrArgGly 435
834 TTGATACGCGCTTTCGACGACGATGCT 863
436 LeuPheSerSerPheSerThrAlaTyrSer 445

```

seq_name: SwissProt_40:MP5A_LOLPR

seq_documentation_block:

```

ID MP5A_LOLPR STANDARD; PRT; 308 AA.
AC Q40240;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Major pollen allergen Lol p 5a precursor (Lol p Va) (Lol p Ib).
GN LOLPB.
OS Lolium perenne (Perennial ryegrass).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```



```

221 AACCGCTTTTGGGCAACGCAAAATGCGTTTGGACTGGCCCCGGG 270
18 pAlaArgGlyAlaGlnGlnArgGlyHisGlnAlaGlyAlaGlyGlnAlaGly 35
271 TTTTCAAAAACCGGCAACGCAATCGAACAACATTTCAACGCGGTACACGG 320
35 laLeuGlnGlyArgGlnAlaGlnArgAlaAspHisAlaHisGlnGlyArg 51
321 CTGGGAACAGCTGACGAGCTTTGGCAACAGGCGGAGGCTGCTTCA 370
52 LeuGly.....GlyGlyGlnAlaGlyAlaAspGlyAlaGln 65
371 TCACGCGGCAATCGGCACTACGA.....TTGGCGCGGA 405
65 saArgAlaAspGlnGlnGlnHisArgAspGlnAlaGlyLeuGln 82
406 CGGTACATGACGACGAGCTTCCGTT
82 euLeuHisGlnArgAspAlaArgGlyGlyArgHisGlyAlaGlyAl 98
432 .....CCAGCT.....GACCGCCATGACA 451
99 ProAspGlyProGlyArgAlaHisGlnGlnGlnArgGlnGln 115
452 AGCCCGCAAAATCAAGCGATAGCAAAATCATGACGCGCGGAGGTG 501
115 uAlaGlyGlnAlaGlyArgGlnGlnLeuGlyAspGlyAspValGly 131
502 CGCGGCAAGCAACCGCCCGCCGCGCATACAGCGGTCACAAACAAT 551
132 .....ArgHisAlaValAlaAspHisAspArgGlyArgGln 144
552 CATCAAGCGCTGCGCGCGCGCGCAACCAT.....584
145 GlnGlnAlaGlnGlyThrGlyAlaGlyGlnArgAlaGlnAspLeuVal 161
585 .CATCTGCGCGCGACAGCTCTTCTCCGCGAGAG.....CGCGCGC 627
161 uArgAlaAlaArgArgGlnArgGlnArgGlnHisLeuAlaHisArg 178
628 GTGCGCGGATTTTGGCAACCGCATACACCATAC.....ACTGCG 674
178 LyAlaGlyGlyArgArgAlaGlyAsnArgGlyGlnAspGlyAla 194
675 GCGCAAAATGCGACAGCTCAAGCGCTGAAGCCGTTTCTGCTGCG 724
195 AspAspValGlyAlaGlnGlnProAlaGlyGlnProVal.....207
725 AACGCTGCGCGCGAGCAAGGCTTCTG.....752
208 .....GlnProArgArgGlnAlaLeuGlnValLeuGlnAlaGly 223
753 .....GTTGCATCCGCGCGCGCGCAAGGGAATGAGAGGAA 791
223 laGlnGlnAspLeuAlaHisProGlnGln.....GlnArgGln 235
792 CAAGCGCGCAGA...TGCGCGCGGTTCACCG 821
236 GlyGlyGlnArgProAlaArgGlyAlaPro 246
seq_name: SwissProt_40:VE2_HPVA7
seq_documentation_block:
ID VE2_HPVA7 STANDARD; PRT; 506 AA.
AC P22420;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Regulatory protein E2.
GN E2.
OS Human papillomavirus type 47.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
Papillomavirus.

```

```

OX NCB1_TaxID=10594;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90281611; PubMed=2162112;
RA Kiyono T., Adachi A., Ishibashi M.;
RT "Genome organization and taxonomic position of human papillomavirus
RL type 47 inferred from its DNA sequence.";
RL Virology 177:401-405(1990).
CC -! FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACNNNNNGGT-3') PRESENT
CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC BY SPERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC REPLICATION.
CC -! SUBUNIT: BINDS DNA AS A DIMER.
CC -! SUBCELLULAR LOCATION: Nuclear.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: M32305; AAA66979.1;
DR PIR: D35324; W2M147.
DR HSSP: P03122; 280p.
DR InterPro: IPR001866; E2_N.
DR InterPro: IPR000427; Early2_C.
DR Pfam: PF00511; E2_C; 1.
DR Pfam: PF00508; E2_N; 1.
DR ProDom: PD000672; Early2_C; 1.
DR ProDom: PD000678; E2_N; 1.
DR Early protein: Transcription regulation; Activator; DNA-binding;
DR Trans-acting factor; DNA replication; Repressor; Nuclear protein.
DR Trans-acting factor; DNA replication; Repressor; Nuclear protein.
KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
SQ SEQUENCE 506 AA; 57478 MW; 92C37F4BFB725065 CRC64;

alignment_scores:
Quality: 108.00 Length: 267
Ratio: 0.857 Gaps: 13
Percent Similarity: 47.191 Percent Identity: 23.970

alignment_block:
US-09-303-518D-571 x VE2_HPVA7
Align seg 1/1 to: VE2_HPVA7 from: 1 to: 506

44 CCATGCATCCGTTGACCGCGCTGCTCA.....AATGC 78
111 ||||| ||||| |||||
197 ProValThrSerThrProGlySerProGlyGlnThrAspPr 213
79 CTCTCCGCTGCTGCTTCTGTCGACACAGCGTGAACCGCGTGG 128
213 oAspThrSerSerIysThrProThrThrThrAlaAlaThrAspThr 230
129 ACATCTGGGCTTTTACCTTTAAAGAAAGACCGCGCGCATGCTGCA 178
230 erProArgArgGlnSerIleAsnIysSerIleGlnThrGluThrLys 246
179 ATA.....TGCGCGAGCGGTTGAACCGCGACAGCGAGCGTCAAA 222
247 ArgArgGlyTyrGlyArgArg.....ProSerSerArgThrArg 260
223 GCGGTTTGGGAACGCAAAATGCGGTTTGAACCTGCCCGCGCTT 272
260 gProGlnThrHisGlnArg..... 267
273 TTTCAAAAACCGGCAAGACATCGAACAATGTCACAGCGGTACAGCGCT 322
||||| ||||| |||||

```



```

295 lnglnleuglynglynglnserglytyrtyrprothserproglngln 311
435 CCTGACCGCCATGTACAGCCGCCGAAATCAACGATAGCAAAATCA 484
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
312 Proglynglynglynglnproglynglynglnproalnglnlygl 328
485 TGCAGCGCGGACAGGTGCGCGCAAGCAAGCAAGCCGCCACCGGCTA 534
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
328 nglnglynglynglynglynglynglynglynglynglynglyng 343
535 CAAGGGGTCAACAATCATCATCAAGC.....CCTGCGCGCGCGGAGGC 578
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
343 lnglnlynglnlnproglynglynglnlnproglynglynglnpro 359
579 AACCATCATCTGCGCCGACACGTCCTCTCCGAGAGCAAGCGCGGCG 628
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
360 glytyrtyrprothserproglnglnserglynglynglynglyng 373
629 TGTGGCGGATTTTTCGCAACCTGCATACACCATGACCT..... 671
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
374 ....Proglytyrtyr...ProthserSerglnlnprothchlnserg 388
672 .....GGCGCAAAATTTGGCAACGTCGCAAGG 698
388 lnglnproglynglynglynglynglynglynglynglynglyngln 404
699 CGTGAACACCT 710
: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
405 AlaInglnlnpro 408

```

seq_name: SwissProt_40:SFR2_MOUSE

```

seq_documentation_block:
ID SFR2_MOUSE STANDARD: PRT: 221 AA.
AC 062093;
DT 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Splicing factor, arginine/serine-rich 2 (Splicing factor SC35) (SC-35)
DE (Splicing component, 35 kDa) (PR264 protein).
GN SFRS2 OR PR264.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98447613; PubMed=9774382;
RA Yang L., Embree L.J., Tsai S., Hickstein D.D.;
RT "Oncoprotein TLS interacts with serine-arginine proteins involved in
RT RNA splicing."
RL J. Biol. Chem. 273:27761-27764(1998).
RL [2]
RP SEQUENCE OF 1-121 FROM N.A.
RA STRAIN=129/SV; TISSUE=Liver;
RA Galliard C., Perbal B.;
RA Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: NECESSARY FOR THE SPLICING OF PRE-mRNA. IT IS REQUIRED
CC FOR FORMATION OF THE EARLIEST ATP-DEPENDENT SPLICING COMPLEX AND
CC INTERACTS WITH SPICEOSOMAL COMPONENTS BOUND TO BOTH THE 5' AND
CC 3' SPLICING SITES DURING SPLICOSOME ASSEMBLY. IT ALSO IS REQUIRED
CC FOR ATP-DEPENDENT INTERACTIONS OF BOTH U1 AND U2 SNRNPS WITH PRE-
CC mRNA (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE RS
CC DOMAIN (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RNM).
CC -1- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.
CC .....
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF077858; AAC1000.1; -.
CC DR EMBL: X98511; CA67134.1; -.
CC DR HSSP: P09651; IUP1.
CC DR MGD: MGI:98284; Str2.
CC DR InterPro: IPR00504; RRM.
CC DR Pfam: PF00076; rrm; 1.
CC DR SMART: SM00360; RRM; 1.
CC DR PROSITE: PS50102; RRM; 1.
CC DR PROSITE: PS00030; RRM_RNP_1; 1.
CC KW Nuclear protein; RNA-binding; mRNA splicing; Phosphorylation.
CC FT DOMAIN 14 92 RNA-BINDING (RRM).
CC FT DOMAIN 116 116 GLY-RICH (HINGE REGION).
CC FT DOMAIN 117 221 ARG/SER-RICH (RS DOMAIN).
CC SO SEQUENCE 221 AA; 25476 MW; 68121AC4D35714FA CRC64;

```

alignment_scores:

```

Quality: 105.50 Length: 211
Ratio: 1.045 Gaps: 7
Percent Similarity: 47.867 Percent Identity: 22.749

```

alignment_block:

US-09-303-518D-571 x SFR2_MOUSE

Align seg 1/1 to: SFR2_MOUSE from: 1 to: 221

```

258 ACITGGCCCCCGCGCTTTTCAAAAACCGGAGACATGCAACAAATGTCA 307
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29 Thrleuargargvalphegluylstyrglyargvalglalyspvaltyrll 45
308 AACCGGTACACGCGCTGGAAACACGTCGACGAGCTTTGA..... 347
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
45 eproargaspargtyrthlysglnserargglyphealphevalargp 62
347 .....
62 hehisplysargspalagluaspalamelaspalamelaspolyala 78
348 .....CAAGGCGAAGGCGCTGTCATGACGCGGACATCGGACACT 391
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
79 ValLeuaspGLyArgGLuLeuArgValGlnMetAla..... 90
392 AGGATTTGGGCGGACGCTACATCAGCCAGAGCTTCGTCACCTGACC 441
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
91 .ArgtyrGLyArg.....ProProAsps 98
442 GCCATGTACAAAGCCGCAAAATCAAAACGCTATGACAAATCATGACGC 491
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
98 eRhIs...HisSerArgArgGLyProProProArgArgTyrglygly 113
492 GGGCAGGTTGCGCGGCAAAAGGCAAAACCGCCGCGCATCAAGGGG 541
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
114 GLyTyrgLyArgSerArgSerProArgArgArgArgArgSerArgse 130
542 TCAAAACAATCATCAAGCCCTGCGCGCGGCGGAGCAACCATCATCTG 591
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
130 ArgSerArgSerArgSerArgSerArgSerArgSerArgTyrglySerArgS 147
592 CCGCAGCAGTCCCTTCCGAGAGAGCGCGCGGTGGGGCGGATTT 641
: ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
147 eRlySerArgSerArgTyrgThArgSerArgSerArg..... 158
642 TTTCGGCAACCTGCATACACATGACATGCGGCAAAATTTGGACAGC 691
: ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
159 .....
692 TCAAAAGCGTGAACCCGTTTCTGCTGCAAGACCTGCGCCA...C 738
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
167 gArgSerTySerTySerSerSerValSerArgSerArgSerArgSera 184

```

739 GGACAAAGCTTCGTGTCACATCCGCCGCTCCAGGGAATTGAACG 788
|||||
184 rgsrArgrSerArgrSerProProValSer.....LysArg 198
789 CACAAAGCCACGATCCGCCGCTTCACACG 821
|||||
199 GluserLysrArgrSerArgrSerPro 209

seq_name: SwissProt_40:CCAA_RABIT

seq_documentation_block: PRT: 2424 AA.
ID CCAA_RABIT STANDARD:
AC P27884; P27883;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Voltage-dependent P/O-type calcium channel alpha-1A subunit (Calcium channel, L type, alpha-1 polypeptide isoform 4) (Brain calcium channel 1) (B1)
GN CACNA1A OR CACNA1A4 OR CACNA4 OR CACNA3.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91187110; PubMed=1849233;
RA Mori Y., Friedrich T., Kim M.-S., Mikami A., Nakai J., Ruth P., Bosse E., Hofmann F., Flocke V., Flocke T., Mikoshiba K., Imoto K., Tanabe T., Numa S.;
RA Imoto K., Tanabe T., Numa S.;
RT "Primary structure and functional expression from complementary DNA of a brain calcium channel.";
RL Nature 350:398-402(1991).
RN [2]
RP BETA-SUBUNIT BINDING DOMAIN AND MUTAGENESIS.
RX MEDLINE=94150724; PubMed=7509046;
RA Pragnell M., de Waard M., Mori Y., Tanabe T., Sutch T.P., Campbell K.P.;
RT "Calcium channel beta-subunit binds to a conserved motif in the I-II cytoplasmic linker of the alpha 1-subunit.";
RL Nature 368:67-70(1994).
CC -1- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIATE THE ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION, CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1A GIVES RISE TO P AND/OR Q-TYPE CALCIUM CURRENTS. P/O-TYPE CALCIUM CHANNELS BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND ARE BLOCKED BY THE FUNNEL TOXIN (FTX) AND BY THE OMEGA-ACETOXIN-IVA (OMEGA-AGA-IVA). THEY ARE HOWEVER INSENSITIVE TO DIHYDROPYRIDINES (DHP), AND OMEGA-CONOTOXIN-GVIA (OMEGA-CTX-GVIA).
CC -1- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- ALTERNATIVE PRODUCTS: IN THE BRAIN, A SHORT ISOFORM BI-1/A-1 AND A LONG ISOFORM BI-2/A-2 (SHOWN HERE), ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: BRAIN-SPECIFIC. PURKINJE CELLS CONTAIN PREDOMINANTLY P-TYPE VSCC, THE Q-TYPE BEING A PROMINENT CALCIUM CURRENT IN CEREBELLAR GRANULE CELLS.
CC -1- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S4, S6) AND ONE POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.

CC -1- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X57477; CAA40715.1; -
CC EMBL: X57689; CAA40872.1; -
CC EMBL: X57476; CAA40714.1; -
CC EMBL: X57688; CAA40871.1; -
CC InterPro: IPR000637; AT_hook.
CC InterPro: IPR002077; Ca_channel.
CC InterPro: IPR002111; Cal_channel_Tppl.
CC InterPro: IPR000636; Calion_chan_non_lig.
CC InterPro: IPR001682; Channel_pore_Ca_Na.
CC Pfam: PF00520; Ion_trans; 4.
CC PRINTS: PR00167; CCHANNEL.
CC SMART: SM00384; AT_hook; 1.
CC Ion channel; Transmembrane; Ion transport; Voltage-gated channel;
CC Calcium channel; Glycoprotein; Repeat; Multigene family;
CC Calcium-binding; phosphorylation; Alternative splicing.
CC REPEAT 85 363
CC REPEAT 473 717
CC REPEAT 1240 1523
CC REPEAT 1560 1823
CC REPEAT 1 98
CC DOMAIN 99 117
CC DOMAIN 118 135
CC DOMAIN 136 155
CC DOMAIN 156 167
CC DOMAIN 168 185
CC DOMAIN 186 190
CC DOMAIN 191 209
CC DOMAIN 210 228
CC DOMAIN 229 248
CC DOMAIN 249 335
CC DOMAIN 336 360
CC DOMAIN 361 487
CC DOMAIN 488 506
CC DOMAIN 507 521
CC DOMAIN 522 541
CC DOMAIN 542 568
CC DOMAIN 569 578
CC DOMAIN 579 597
CC DOMAIN 598 616
CC DOMAIN 617 636
CC DOMAIN 637 689
CC DOMAIN 690 714
CC DOMAIN 715 1253
CC DOMAIN 1254 1272
CC DOMAIN 1273 1288
CC DOMAIN 1289 1308
CC DOMAIN 1309 1320
CC DOMAIN 1321 1339
CC DOMAIN 1340 1350
CC DOMAIN 1351 1369
CC DOMAIN 1370 1388
CC DOMAIN 1389 1408
CC DOMAIN 1409 1495
CC DOMAIN 1496 1520
CC DOMAIN 1521 1575
CC DOMAIN 1576 1604
CC DOMAIN 1605 1609
CC DOMAIN 1610 1629
CC DOMAIN 1630 1637
CC DOMAIN 1638 1656
CC DOMAIN 1657 1665

| | | | |
|----|----------|----------|------------------------------------|
| FT | TRANSMEM | 1666 | 1684 |
| FT | DOMAIN | 1685 | 1703 |
| FT | TRANSMEM | 1704 | 1723 |
| FT | DOMAIN | 1724 | 1795 |
| FT | TRANSMEM | 1796 | 1820 |
| FT | DOMAIN | 1821 | 2424 |
| FT | DOMAIN | 13 | 18 |
| FT | DOMAIN | 727 | 732 |
| FT | DOMAIN | 1004 | 1010 |
| FT | DOMAIN | 1012 | 1017 |
| FT | DOMAIN | 2219 | 2227 |
| FT | DOMAIN | 2242 | 2246 |
| FT | DOMAIN | 2288 | 2297 |
| FT | DOMAIN | 2298 | 2301 |
| FT | DOMAIN | 2372 | 2377 |
| FT | DOMAIN | 2411 | 2416 |
| FT | DOMAIN | 383 | 400 |
| FT | SITE | 318 | 318 |
| FT | SITE | 668 | 668 |
| FT | SITE | 1469 | 1469 |
| FT | SITE | 1765 | 1765 |
| FT | MOD. RES | 1831 | 1831 |
| FT | CAR.BIND | 1849 | 1860 |
| FT | CAROHD | 283 | 283 |
| FT | CAROHD | 1665 | 1665 |
| FT | VARSPLC | 772 | 1051 |
| FT | VARSPLC | 772 | 1120 |
| FT | VARSPLC | 1857 | 1884 |
| FT | VARSPLC | 2230 | 2273 |
| FT | VARSPLIC | 2274 | 2424 |
| FT | VARIANT | 419 | 419 |
| FT | VARIANT | 877 | 877 |
| FT | VARIANT | 1104 | 1104 |
| FT | MUTAGEN | 386 | 386 |
| FT | MUTAGEN | 389 | 389 |
| FT | MUTAGEN | 392 | 392 |
| FT | MUTAGEN | 400 | 400 |
| SO | SEQUENCE | 2424 AA: | 273228 MM: F7CC4D0AB4B45604 CRC64. |

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 105.50 | Length: 313 |
| Ratio: | 0.776 | Gaps: 19 |
| Percent Similarity: | 43.450 | Percent Identity: 22.364 |

```
alignment_block:
```

```

US-09-303-518D-5/1 x CCAA_RABIT  ..
Align seg 1/1  to: CCAA_RABIT  from: 1  to: 2422

```

[illegible][illegible]

DE (BASS1) (Protein C21orf50).
 OS SON OR NREBP OR DBPS OR C21ORF50 OR KIAA1019.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NX NCBI_TaxId=9606;
 RX [1]
 RP SEQUENCE FROM N.A. (ISOFORM A; B; C; D; E AND F).
 RC MEDLINE=21564202; PubMed=11707072;
 RA Raymond A., Friedl M., Neergaard Henriksen C., Chapot F.,
 RA Deutsch S., Ucla C., Rossier C., Lyle R., Guipponi M.,
 RA Antonarakis S.E.;
 RT "From PREDS and open reading frames to cDNA isolation: revisiting the
 RT Human Chromosome 21 Transcription Map";
 RL Genomics 78:46-54(2001).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM G).
 RC TISSUE=Liver;
 RX MEDLINE=21316479; PubMed=11306577;
 RA Sun C.-T., Lo W.-Y., Wang I.-H., Lo Y.-H., Shou S.-R., Lai C.-K.,
 RA Ting L.-P.;
 RT "Transcription repression of human hepatitis B virus genes by negative
 RT regulatory element-binding protein/SON";
 RL J. Biol. Chem. 276:24059-24067(2001).
 RN [3]
 RP SEQUENCE OF 1-689 FROM N.A. (ISOFORM H).
 RC TISSUE=Placenta;
 RA Casadei R., Strippoli P., D'Addabbo P., Canaider S., Lenzi L.,
 RA Vitale L., Giannone S., Carinci P., Zannotti M.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 1-130 FROM N.A.
 RC TISSUE=Smooth muscle;
 RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
 RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
 RA Yamada K., Fujii Y., Ozaki K., Hirao M., Omori Y., Ota T., Suzuki Y.,
 RA Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
 RA Isogai T., Sugano S.;
 RT "NEBO human cDNA sequencing project";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 1-114 FROM N.A.
 RC TISSUE=Blood;
 RA Ye M., Zhang O.H., Zhou J., Shen Y., Wu X.Y., Guan Z.O., Wang L.,
 RA Fan H.Y., Mao Y.F., Dai M., Huang O.H., Chen S.J., Chen Z.;
 RT "Human partial CDS from cd34+ stem cells";
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE OF 437-2426 FROM N.A. (ISOFORM B).
 RC TISSUE=Brain;
 RX MEDLINE=99397452; PubMed=10470851;
 RA Kikuno R., Nagase T., Ishikawa K.-I., Hirose M., Miyajima N.,
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIV.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro";
 RL DNA Res. 6:197-205(1999).
 RN [7]
 RP SEQUENCE OF 554-2426 FROM N.A. (ISOFORM A).
 RC MEDLINE=92049296; PubMed=1944255;
 RA Chumakov I.M., Berdichevskii F.B., Sokolova N.V., Reznikov M.V.,
 RA Prisojov V.S.;
 RT "Identification of a protein product of a novel human gene SON and
 RT the biological effect upon administering a changed form of this gene
 RT into mammalian cells";
 RL Mol. Biol. (Mosk) 25:731-740(1991).
 RN [8]
 RP SEQUENCE OF 709-1079 FROM N.A. (ISOFORM I).
 RC TISSUE=Placenta;
 RX MEDLINE=93062885; PubMed=1435774;
 RA Bliskovskii V.V., Kirillov A.V., Zakhariev V.M., Chumakov I.M.;
 RT "The human son gene: the large and small transcripts contains various
 RT 5'-terminal sequences";

RL Mol. Biol. (Mosk) 26:807-812(1992).
 RN [9]
 RP SEQUENCE OF 1009-1131 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=93062884; PubMed=1435773;
 RA Bliskovskii V.V., Berdichevskii F.B., Tkachenko A.V., Belova M.E.,
 RA Chumakov I.M.;
 RT "Coding part of the son gene small transcript contains four areas of
 RT complete tandem repeats";
 RL Mol. Biol. (Mosk) 26:793-806(1992).
 RN [10]
 RP SEQUENCE OF 1145-2426 FROM N.A. (ISOFORM F).
 RX MEDLINE=93048367; PubMed=1424986;
 RA Mationi T., Hume C.R., Konigorski S., Hayes P., Osterweil Z.,
 RA Lee J.S.;
 RT "A cDNA clone for a novel nuclear protein with DNA binding
 RT activity";
 RL Chromosoma 101:618-624(1992).
 RN [11]
 RP SEQUENCE OF 1692-2175 FROM N.A. (ISOFORM A).
 RX MEDLINE=99039788; PubMed=1054499;
 RA Berdichevskii F.B., Chumakov I.M., Kiselev L.L.;
 RT "Decoding of the primary structure of the son3 region in human
 RT genome: identification of a new protein with unusual structure and
 RT homology with DNA-binding proteins";
 RL Mol. Biol. (Mosk) 22:794-801(1988).
 RN [12]
 RP SEQUENCE OF 1939-2426 FROM N.A. (ISOFORM J).
 RC TISSUE=Cerebellum;
 RX MEDLINE=99439804; PubMed=10509013;
 RA Greenhalf W., Lee J., Chaudhuri B.;
 RT "A selection system for human apoptosis inhibitors using yeast";
 RL Yeast 15:1307-1321(1999).
 CC -1- FUNCTION: Represses hepatitis B virus (HBV) core promoter activity
 CC and transcription of HBV genes and production of HBV virions.
 CC Binds to the consensus DNA sequence: 5'-GAG(G)AAG(C)AG(C)CC-3'.
 CC Might protect cells from apoptosis. Might be involved in pre-mRNA
 CC splicing (by similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear with a speckled distribution.
 CC -1- ALTERNATIVE PRODUCTS: 10 isoforms: A, B, C, D, E, F (shown here),
 CC G, H, I and J; may be produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Widely expressed, with the higher expression
 CC seen in leukocyte and heart.
 CC -1- DOMAIN: Contains 8 types of repeats which are distributed in 3
 CC regions.
 CC -1- MISCELLANEOUS: Colocalizes with the pre-mRNA splicing factor
 CC SF2/SC-35.
 CC -1- SIMILARITY: CONTAINS 1 G-PATCH DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 DBRM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.
 CC -1- CAUTION: ISOFORM A SEQUENCE FROM REF.7 DIFFERS FROM THAT SHOWN
 CC DUE TO A FRAMESHIFT.
 CC -1- CAUTION: ISOFORM F SEQUENCE FROM REF.10 DIFFERS FROM THAT SHOWN
 CC DUE TO A FRAMESHIFT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AF380179; AL34497.1; -
 DR EMBL; X63753; CAA45282.1; ALU_FRAME.
 DR EMBL; M36428; AAA36624.1; -
 DR EMBL; AF380180; AL34498.1; -
 DR EMBL; AF380181; AL34499.1; -
 DR EMBL; AF380182; AL34500.1; -
 DR EMBL; AF380183; AL34501.1; -
 DR EMBL; AF380184; AL34502.1; -
 DR EMBL; AY026895; AAK07692.1; -
 DR EMBL; AF435977; AL30810.1; -
 DR EMBL; X63751; CAC69885.1; -


```

DR EMBL: AB028942; BA028971.1; -
DR EMBL: X63071; CA044793.1; ALT_FRAME.
DR EMBL: AF139897; AAD50078.1; -
DR EMBL: S47238; AAB23945.1; -
DR EMBL: AK024752; BAB14985.1; -
DR EMBL: AF161428; AAF28888.1; -
DR EMBL: AF161430; AAF28990.1; -
DR PIR: P00099; P00099.
DR MIM: 182465; -
DR InterPro: IPR001159; DS_RBD.
DR InterPro: IPR000467; G_patch.
DR Pfam: PF00035; dsrm; 1.
DR Pfam: PF01585; G_patch; 1.
DR SMART: SM00358; DSRM; 1.
DR SMART: SM00443; G_patch; 1.
DR PROSITE: PS50137; DS_RBD; 1.
DR PROSITE: PS50174; G_PATCH; 1.
KW RNA-binding; DNA-binding; Nuclear protein; Repeat;
KW Alternative splicing.
FT DOMAIN 726 895 17 X 10 AA TANDEN REPEATS OF L-A-[ST]-
[NSG]-[TS]-MDSOM.
FT 912 988 11 X 7 AA TANDEN REPEATS OF [DR]-P-Y-R-
[LI][AG][OHP].
FT 1006 1126 14 X 6 AA REPEATS OF [ED]-R-S-M-M-S.
FT REPEAT 1006 1126 1-1.
FT REPEAT 1014 1019 1-2.
FT REPEAT 1021 1026 1-3.
FT REPEAT 1030 1035 1-4.
FT REPEAT 1038 1043 1-5.
FT REPEAT 1046 1051 1-6.
FT REPEAT 1055 1060 1-7.
FT REPEAT 1063 1068 1-8.
FT REPEAT 1071 1076 1-9.
FT REPEAT 1080 1085 1-10.
FT REPEAT 1089 1094 1-11.
FT REPEAT 1100 1105 1-12.
FT REPEAT 1111 1116 1-13.
FT REPEAT 1121 1126 1-14.
FT DOMAIN 1147 1179 3 X 11 AA TANDEN REPEATS OF P-P-L-P-P-E-E-
P-P-[TME]-[MTG].
FT 1359 1390 4 X 8 AA TANDEN REPEATS OF V-L-E-SS-
[AVT]-VT.
FT DOMAIN 1925 1994 7 X 7 AA REPEATS OF P-S-R-R-S-R-[TS].
FT REPEAT 1925 1931 2-1.
FT REPEAT 1933 1939 2-2.
FT REPEAT 1960 1966 2-3.

alignment_scores:
Quality: 105.50 Length: 228
Ratio: 1.088 Gaps: 10
Percent Similarity: 42.544 Percent Identity: 27.632

alignment_block:
US-09-303-518D-571 x SON_HUMAN
Align seg 1/1 to: SON_HUMAN from: 1 to: 2426

152 AGGAAGACCGCGCGCGCATCGTCGCAATATGCGGCGGTTTGAAAC 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1871 ArgArgSerSerArgSerArgSerArgSerArgSerArgSerValSe 1887
202 CCGACACGCGGACGCGTCACAAAGCGTTTGGCGAAACGCAAAATGCG 251
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1887 rlygglulysarglylsargserPro..... 1896
252 TTGGAACTTGCCCGCGCTTTTCAAAAAACCGAAGACATGAAACAA 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1896 .....LysHisArgSerLysSerArgGlu 1903
302 TGTTCAGAGCGGTACACGCTGGGACACATGACGCGCTTTGACACAG 351
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1904 ...ArgLysArgLysArgSerSerArgAspAsnArgLysThrValAr 1919

```

```

352 GCGCAAGGCGTGTCTTCATCAGCGCGACATCGGACATGACATTGGG 401
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1919 galArgSerArgThrProSerArgSerArgSerArgSerHisThrProSe 1936
402 CCGACGCTCATCTACGCGACGAGCTTCGCT.....TCACCTGA 439
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1936 rgaArgArgSerArgSerValGlyArgArgSerProSerHisSer 1952
440 CCGCATGTATACAGCGCGCAAAATCAACCGATACCAAAATCATGACG 489
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1953 ProSerArgSerArgThrProSerArgSerArgThrProSerAr 1969
490 GCGGCGAGGTGGCGCGCAAAAGCAAAACCGCCACCGCATACAAAG 539
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1969 gArgSer.....ArgThrPro..... 1974
540 GGTCAAAATATCATACAGCCCTGGCGCGCGGCGGACACATCATCC 589
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1975 ..SerArgSerArgThrProSerArgSerArgThrProSerArg 1990
590 TGCCCGACACGTCCTCTCCGACGAGCGCGCGCGTGGCGGAT 639
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1991 ArgSerArgThrProSerArgArgArgArgSerArgVal..... 2004
640 TTTTGGCAAACTCATACACATGACATGCGCGCAAAATGCGACA 689
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2005 .....ValArgArg 2007
690 CGTCAAGGCGTGAANACCTGTTTTCG.....CGGACGCGCTGC 733
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2008 ArgSerPheSerHisSerProValArgLeuArgSerArgThrProLe 2024
734 CCGACGACACAGCTTCGTTGACATCCGCC.....CGTC 771
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2024 uArgArg.....ArgPheSerArgSerProLeargArgLysArgS 2038
772 CAAGGGAATGACGCGACCAAGCCACGCA 803
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2038 eArgSerSerGluArgLysArgSerProLys 2048

seq_name: SwissProt.40:YBPH_APLCA
seq_documentation_block:
ID YBPH_APLCA STANDARD; PRT; 253 AA.
AC P41824;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Y-box factor homolog (APY1).
OS Aplysia californica (California sea hare).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspiidea;
OC Aplousiidae; Aplysia.
OX NCBI_TaxId=6500;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94333813; PubMed=7520012;
RA Skehel P.A., Bartsch D.;
RT "Characterization of a Y-Box factor from Aplysia californica.";
RL Gene 145:231-235(1994)
CC 1- FUNCTION: BINDS RNA IN VITRO.
CC 1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC 1- PTM: PHOSPHORYLATION ACTIVATES IN VITRO RNA-BINDING
CC (BY SIMILARITY).
CC 1- SIMILARITY: BELONGS TO THE COLD-SHOCK DOMAIN (CSD) FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

DR EMBL: U02684; AAA60373.1; -
 DR HSP; P41016; ICGO.
 DR InterPro: IPR002059; Cold_shock.
 DR Pfam: PF00313; CSD; 1.
 DR PRINTS: PR00050; COLDSHOCK.
 DR PRODOM: PD000621; Cold_shock; 1.
 DR SMART: SM00357; CSD; 1.
 DR PROSITE: PS00352; COLD_SHOCK; 1.
 DR Transcription regulation; RNA-binding; Nuclear protein;
 KW Phosphorylation.
 FT DOMAIN 35 99
 FT MOD_RES 226 226 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
 FT DOMAIN 116 138 ARG-RICH REGION (BASIC).
 FT DOMAIN 152 173 ARG-RICH REGION (BASIC).
 FT DOMAIN 186 204 ARG-RICH REGION (BASIC).
 FT DOMAIN 234 253 ARG-RICH REGION (BASIC).
 FT SEQUENCE 253 AA; 29439 MW; AA3400013FEACEE3 CRC64;

alignment_scores:
 Quality: 105.00 Length: 169
 Ratio: 1.265 Gaps: 10
 Percent Similarity: 49.112 Percent Identity: 28.402

alignment_block:
 US-09-303-518D-571 x YBFH_APLCA

Align seg 1/1 to: YBFH_APLCA from: 1 to: 253

420 GCACCTTCCTT.....CCACCTGACCGCATGTACAGCGCCGCAAAA 463
 |||||:|||||
 96 Alalalsnvalthrglprogluglyserasvalglnglysetlystys 112
 :|||||:|||||
 464 TCAAGCGATGACAAAT.....C 483
 :|||||:|||||
 112 ralaalalspargrargrpheargrarglyglytrpyrproargphea 129
 :|||||:|||||
 484 ATGACGCGGCGAGGTCGCCGCAAGCAAGCAAGCGCCGCGCATG 533
 :|||||:|||||
 129 rgllyglylgrarglylgrprotrrglinspmet..... 141
 :|||||:|||||
 534 ACAAGGGGTCAACCAATCAAGCGCGCGCGCGCGAGGCAACCA 583
 :|||||:|||||
 142 AspspjlalaprproaspheleProserProarglylgrg..... 156
 :|||||:|||||
 584 TCATCCTCGCGACCGACCTCCTCCTCGCGAGGAGCG..... 623
 :|||||:|||||
 157Argllyargprolyrtyrgrlnasargargtyrphelgyp 170
 :|||||:|||||
 624CGCGCTGTGGCGGATTTTTCGGCAACCTGC..... 656
 :|||||:|||||
 170 roproargrarglygly.....Arglnrtyrleuglglugly 184
 :|||||:|||||
 657ATACACCATGACACGCGGCAAAATTTGGCAAGCTCAAG 697
 :|||||:|||||
 185 Glutyrglnleuglnargaspelnglyphearg.....Glyalargargpr 200
 :|||||:|||||
 698 GGGTGAACCCCTGT.....T 714
 :|||||:|||||
 200 opherlyrargproleuleuargthrtThrserlnglyleuleuargargr 217
 :|||||:|||||
 715 TTCTGTGCGAAGCGCTGCGCGACGACAGGCTTCGTTGCGACATCG 764
 :|||||:|||||
 217 rpleuleuargleuargrargrthrtglngrlyargthrserlngla 233
 :|||||:|||||
 765CCCCGCCAAGGGGAATGAACGCGCAACCAAGCCACGATG 805
 :|||||:|||||
 234 Argargarglgrargprotrpelyleuargproglnglnargprplyspr 250
 :|||||:|||||
 806 CGCGCGT 812
 :|||||:|||||
 250 oarggin 252

seq_name: SwissProt_40:CYF_CHLVU

seq_documentation_block:
 ID CYF_CHLVU STANDARD; PRT; 341 AA.
 AC P56316;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Apocytochrome F precursor.
 GN PETA.
 OS Chlorella vulgaris.
 OC Chloroplast.
 CC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
 CC Chlorellaceae; Chlorella.
 NCBI_TaxID=3077;
 RX MEDLINE=97303241; PubMed=9159184;
 RA Wakasugi T., Nagai T., Kapoor M., Tsudzuki T., Suzuki Y., Hamada A., Ohta T.,
 Tsudzuki J., Nakashima K., Tsudzuki M., Sugita M., Ito M., Ito S.,
 Inamura A., Yoshinaga K., Sugita M.;
 "Complete nucleotide sequence of the chloroplast genome from the
 green alga Chlorella vulgaris: the existence of genes possibly
 involved in chloroplast division".
 RT Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972(1997).
 RL -1- FUNCTION: TRANSLOCATES PROTONS ACROSS THE THYLAKOID MEMBRANE AND
 TRANSFERS ELECTRONS FROM THE RIESKE IRON-SULFUR PROTEIN AND PASSES
 THEM TO PLASTOCYANIN; THIS FUNCTION IS VERY SIMILAR TO THAT OF
 MITOCHONDRIAL CYTOCHROME C1.
 CC -1- SUBUNIT: MEMBER OF THE CYTOCHROME B6/F COMPLEX INCLUDING
 CYTOCHROME B6, CYTOCHROME F AND PROBABLY AN IRON SULFUR PROTEIN.
 CC -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane (Probable).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AB001684; BAA57987.1; -
 DR HSP; P36438; 1HCZ.
 DR InterPro: IPR002325; Apocyl-F.
 DR InterPro: IPR000345; Cytochrome_bnd.
 DR Pfam: PF01333; Apocytochrome_F; 1.
 DR PRINTS: PR00610; CYTOCHROME_F; 1.
 DR PROSITE: PS00190; CYTOCHROME_C; 1.
 DR Electron transport; Heme; Chloroplast; Thylakoid;
 KW Photosynthesis; Photosystem I; Photosystem II; Transit peptide;
 KW Transmembrane.
 FT TRANSIT 1 56
 FT CHAIN 57 341
 FT METAL 57 57
 FT BINDING 77 77
 FT BINDING 80 80
 FT METAL 81 81
 FT TRANSMEM 307 327
 FT SEQUENCE 341 AA; 37350 MW; FBC53FC5D09465F CRC64;

alignment_scores:
 Quality: 105.00 Length: 311
 Ratio: 0.745 Gaps: 13
 Percent Similarity: 45.338 Percent Identity: 21.543

alignment_block:

US-09-303-518D-571 x CYF_CHLVU

Align seg 1/1 to: CYF_CHLVU from: 1 to: 341

```

7  CGTTTAAATTCAGGCTGTTTCCCTTTCGCAACCGCATGCATACCT 56
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6  LysLeuGluPheAsnPheIleProAsnLeuLys.....LysHisAla 20
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
57 GTTGACCGCCGCTGCATAATGCTCTCCCTGCTGCTGCTCTGCTGCTGC 106
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
20 LpheSerPheTrpGlyGlnAsnGlnAlaIleLeuLysPheSerThrLeu 37
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
107 ACACGCTGGAAACCGCTCGACATCTGGGCTTTTACTTTTAAAGAA 156
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
37 alSerLysGlyAlaLeuValLeuValLysSerPhePheLeuThrAlaSer 53
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
157 GACCCGCGCCGATCTC.....GCCAATATGCGCA 188
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
54 SerAsnAlaTyrProIlePheAlaGlnGlnAsnTyrAlaAsnProArg 70
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
189 GGCG..... 192
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
70 uAlaAsnGlyArgIleValCysAlaAsnCysHisLeuAlaGluProI 87
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
193 .....GTTTGAACCCGACAG..... 210
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
87 IeGluIleGluValProGlnAlaValLeuProAspThrValPheGluAla 103
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
211 .....CAGACGCTCAAAGCGTTTTCGGAAC 239
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
104 ValValLysIleProTyrAspLysGlnIleLysGlnValLeuAlaAsn 120
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
240 GCAAAA.....TGGGTT 253
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
120 YLysLysGlyAspLeuAsnValGlyAlaValLeuIleLeuProAspGly 137
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
254 TGGAACTTGGCCCGGCTTTTCAAAAACCGGAGACATCGAACAATG 303
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
137 heGluIleAlaPro.....ProAspArgIleProGluGlu 148
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
304 TTCAAAACCGGTACACGCTGGGACACGTCGACGCGCTTTGGACAAGG 353
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
149 MetLysAlaLysValGlyLysLeuTyrPheGlnProTyrSerAlaGlu 165
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
354 CGAAGGCTGCTGTTATCATCGCCGACATCGACATGATGATGGGCG 403
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
165 sLysThrIlePheValValGlyProValProGlyLys..... 177
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
404 GACGCTACATCAGCCAGCTTCGTTCCACTGACCCGCTGATGACAAG 453
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
178 ..LysTyrSerGluMetValPheProIle.....LeuSerProasp 190
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
454 CCGCCGAAAATCAAGCGATAGCAAAATC.....ATGCAGGCGG 494
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
191 ProAlaLysThrLysSerIleSerTyrLeuLysTyrProIleTyrAla 207
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
495 CAGGCTGCGCGCAAGGCAAAACCGCCGACCGCATACAGGGGTCA 544
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207 yGlyAsnArgGlyArgGlyGlnValTyrProAspLysSerLysSerAsn 224
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
545 AACCAATATCAAGGCTGCGCGCGCGGCGAGGACATCATCTGCTGCC 594
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
224 sThrIlePheThrAlaSerAlaIleGlyLysIleThrAlaIle..... 238
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
595 GACGCTGCTCCTTCGCGAGGAGGCGGCGCTGTGGCGGATTTT 644
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
239 .....GluProAlaGlyLysLysGlyGly..... 246
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
645 CGGCAACCTCGATACACCATGACATCGGCGGCAAAATGCAACAGTCA 694
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
247 .....TyrThrLeuThrIle.....GluThrAlaAsnGly 257
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
695 AAGCGTGAATAACCTGTTTTCGCTGCGAAGCGCTGCGCGAGCAAA 744
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
257 InsThrIleSer.....GluLysLeuProGlyPro 267

```

```

745 GGCTTGCTGTGCATCCGCCCGCTCCAGG 777
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
268 GluLeuValValAsnIleGlyAspIleValGly 278

```

seq_name: SwissProt.40:5E5_RAT

seq_documentation_block:

```

ID      5E5_RAT      STANDARD;      PRT;      825 AA.
AC      063003;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      5E5 antigen.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Wistar; TISSUE=Brain;
RX      MEDLINE=96015159; PubMed=8537300;
RA      Suzuki E., Kojima N., Yoshimura K., Uyemura K., Ohta K., Akagawa K.;
RT      "Cloning and sequence analysis of cDNA for a possible DNA-binding
RT      protein 5E5 in the nervous system.";
RL      J. Biochem. 118:122-128(1995).
CC      FUNCTION: MIGHT HAVE DNA-BINDING ABILITY.
CC      SUBCELLULAR LOCATION: Nuclear.
CC      TISSUE SPECIFICITY: EXPRESSED IN NEURONS.

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: D37934; BAA07153.1; -
KW DNA-binding; Nuclear protein; Antigen.
SQ SEQUENCE 825 AA; 86831 MW; AF667FE2FD55BDF CRC64;

alignment_scores:

```

Quality: 105.00      Length: 309
Ratio: 0.854      Gaps: 13
Percent Similarity: 39.806      Percent Identity: 23.625

```

alignment_block:

```

US-09-303-518D-571 x 5E5_RAT
Align seg 1/1 to: 5E5_RAT from: 1 to: 825

```

```

6  TCGTTTAAATTCAGGCTGTTTCCCTTTCGCAACCGCATGCATCC 55
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
535 SerGlnThrLeuProAlaLeuAlaGlyAlaProThrAlaHisAla 551
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
56 TGTTCACCGCCCTGCTCAATGCTTCCTGCTGCTGCTTCTGCTG 105
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
551 ValProGlyPro..... 555
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
106 CACACGCTGGGAAACCGGCTGACATCTGCGCTTTTACCTTTAAAGA 155
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
556 .....GlyProAlaAlaAlaThrLeuGlyArgGlyArgGly 569
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
156 AGACCGCGCGGCGATCTGCGCAATATGCGGCGGCTTTAAACCCG 205
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
570 SerThrArgGlyArgGlyArgGlyGlyAlaGly..... 581
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206 ACACGCGAGGTCAAAGCGTTTTCGCGAAGCAAAATGCGGTTG 255
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
582 ....AlaSerLysGlyArg...GlyGlyArgGlyArgGlyArg.... 594
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
256 GAACCTGCCCCGCTTTTCAAAAACCGGAAGACATGCAAAATGTT 305

```

```

595 .....CTGGAACACGTCAGCAGGCTTTGGACA 349
306 CAAAGCGGTACACG.....CTGGAACACGTCAGCAGGCTTTGGACA 349
602 LeuSerGlyThrArgGluAspAlaGlySerProSerAlaArgArgGly 618
350 AGGCGAAGGCGTGTTCATCAACGCCGACATCGGACGATGATTTG 359
618 uGlnArgArgGly..... 623
400 GCGGACGCTACATCAGCAGCAGCTTCGTTCCATCCTGACGCCATGTA 449
624 .....HisGlyProProAlaAla..... 629
450 CAAAGCGGTACACGTCAGCAGGCTTTGGACA 499
630 .....GlyAlaAlaGlnValSerThrArgGlyArgAlaArgGlyGln 645
500 TGGC.....CGCAAAAGCAAAAC 519
645 gThrGlyGlnGluAlaGlnAspGlyLeuLeuProArgGlyArgAspArgL 662
520 GCGCCACCGGCTACAGGCGTCAACAAATCATCAAGGCCCTCGCGC 569
662 euProLeuArgProGlyAspSerAsnGlnArgValGlnArgProGlyHis 678
570 GGGGAGGCAACCAT..... 584
679 ProArgGlyGlyHisGlyAlaIleAsnAlaProSerAlaProAspAla 695
585 .CATCTGCCCGCAGCTCCCTTCCTCCGAGCAAGCGCGCGCGTG. 632
695 rProProHisProArgGlyArgValSerGlnGlnArgGlnArgLeuT 712
633 .....GGCGGATTT..... 641
712 rPArgGlnPheArgValGlyGlyPheProProProProProThArg 728
642 .....TTTGGCAAACTGCATACACATGA 667
729 ProProValLeuLeuProLeuArgLeuThrCysAlaGlyAspP 745
668 CACGCGCGCAAAATGCGACACGTCGAAGCGTGAACCCCTGTTTTC 717
745 oGlyAlaSerArgProGlySerArgG..... 754
718 TGCTGCGAAGCGCTGCCGAGCAAGGCTTCGTTGACATCGGCC. 767
755 .....ProAlaArgArgProArgGlyGlnLeuThr...ProGln 766
768 CGTCCAGGGAATGAGAGGCAACAA 794
767 ArgProSerProPheAlaProGlnGlu 775
seq_name: SwissProt_40:REPT_MOUSE
seq_documentation_block:
ID REPT_MOUSE STANDARD; PRT; 1130 AA.
AC P97347;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Repetlin.
GN RPTN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NMRI; TISSUE=SKIN;
RX MEDLINE=97422611; PubMed=9268637;

```

```

RA Kriegl P., Schuppler M., Koesters R., Mincheva A., Lichter P.,
RA Marks F.:
RT "Repetlin (Rptn), a new member of the 'fused gene' subgroup within the
RT S100 gene family encoding a murine epidermal differentiation
RT protein."
RT Genomics 43:339-348(1997).
CC -1- FUNCTION: NOVEL POTENTIAL PRECURSOR PROTEIN OF THE CORNIFIED CELL
CC ENVELOPE.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-stb.ch).
CC -----
DR EMBL; X99251; CAA67624.1; -.
DR HSSP; P02633; 3ICB.
DR MGD; MGI:1099055; Rptn.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR001751; S100_CABP.
DR Pfam; PF00036; ehand; 1.
DR Pfam; PF01023; S_100; 1.
DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PS00303; S100_CABP; 1.
DR Repeat; Calcium-binding.
KW Repeat; Calcium-binding.
FT DOMAIN 1 91 S-100 LIKE.
FT CA_BIND 22 33 EF-HAND 1 (HIGH AFFINITY) (POTENTIAL).
FT CA_BIND 62 73 EF-HAND 2 (LOW AFFINITY) (POTENTIAL).
FT SEQUENCE 1130 AA; 129884 MW; 7D97BDD25151918D CMC64;

```

alignment_scores: Quality: 105.00 Length: 269
Ratio: 0.814 Gaps: 12
Percent Similarity: 47.955 Percent Identity: 23.420

alignment_block:
US-09-303-518D-571 x REPT_MOUSE
Align seg 1/1 to: REPT_MOUSE from: 1 to: 1130

```

99 CTGTCTGCACACGCTGGGAACCGGCTCG.....ACATCTGCGCTTT 142
||||| ||| |||||: ||| |||
675 LeuSerProHisGlnGlyGlnGlyArgGlnAspGlnSerProHisLe 691
143 ACCTTTTAAAGAGAACCGCGCGCGCATCTCGCAATATCGCGGACGCG 192
| ||||| |||
691 uGlnGlnGlyGlnArgHisAspGlnSerProHisGlnGlnGlyGlyA 708
| ||||| |||
193 GGTGTGAACCCCGACACCGACGCTCA.....AGC 224
|||
708 rGnHisAspGlnSerProHisGlnGlnGlyGlnGlyArgGlnAspLeuSer 724
|||
225 CGTTTTCGGGAACGGCAAAATCGGTTTGAACCTTGCCTCCCGCGCTTT 274
||||| ||| |||
725 SerHisGlnGlyGln...LysGlyArgGlnAspGlnSerProHisLeuG 740
||||| ||| |||
275 TCAAAAACCGGAGAGACATCGAACAATGTTCAAAAGCGGTACACGCTG 324
||||| ||||| |||||: ||| |||
740 yGlnGly...GlyArgHisAspGlnSerProHisArgGlyGlnGly...G 755
|||
325 GAACAGCTGCAGCAGGCTTTGGACAAAGCGGCGGCTGTTCATCAC 374
|||
375 GCGGACATCGGACGCTACGATTTGGCGGCGACGCTCATCATCAGCAGCAGC 424
||||| ||| |||||: ||| |||
772 SerSerHisGlnGlnGlnArgGlnGlnGlnGlnGlnGlnGlnSer..... 785

```

```

425 TTCGGTTCACACGACGCGCATGTACACGCGCCGCAAAATCAACCGATA 474
786 ..... HistPrisrGthrAsparGlnGlnGlnSerPhe 798
475 GACAAATCATATGCGAGCGGCGAGGTCGCGCAAGCAAAACCGCCGCC 524
798 IstyrgLglnThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 814
525 ..... CACCGG...CATACAAGGCGTCA 544
815 AspSerGlnGlnGlnAsnSerGlnThrPheAspThrAspSerGlnGln 831
545 AACAAATCAT.....CAAGCGCTCGCCGCGGCGGAGCAACCATCATC 588
831 nserPheHisPheAspGlnAlaGlnGlnGlnGlnGlnGlnGlnGln 848
589 CTGCCCGACACGTCCTCTCCGAGAGAG.....CGCGCGCGTGTG 632
848 lylGlnThrAsparGlnSerGlnSerHisGlyGlnSerGlnGln 864
633 GCGGATTTTTCGCAAACTGCATACCATGACACTGCG..... 674
865 GlyLysThrGlnAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 881
675 ..... GCGAAATATG 684
881 gThrArgArgAspSerTyValGlnHisSerGlyArgSerGlyLys 898
685 GCACACGTCGCAAGCGGTGAAACCCCTGTTTCTGTCGCAACGCGCTGC 734
898 ergLglnHisSerArgGlnGln.....Val 906
735 CGACGCGCAAGGCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 784
907 ArgGlnGlnSerGlnArgSerHisAspArgArgGlnGlnGlnGln 923
785 ACGGCAA 791
923 nGlnGln 925

```

seq_name: SwissProt_40:YT44_STRFR

```

seq_documentation_block:
ID YT44_STRFR STANDARD; PRT: 395 AA.
AC P20188;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
OS Hypothetical 44.4 kDa protein in transposon TM4556.
OC Streptomyces fradiae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_Taxid=1906;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-TM4556;
RX MEDLINE=90185236; PubMed=2155856;
RA Siemieniak D.R., Slightom J.L., Chung S.T.;
RT "Nucleotide sequence of Streptomyces fradiae transposable element
RL Tm4556, a class-II transposon related to Tn3."
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL: M29297; AAA8562.1; -
DR PTR; J00430; J00430.

```

KW Hypothetical protein; Transposable element.
SQ SEQUENCE 395 AA; 44379 MW; C50B556F0E3B2838 CRC64;

alignment_scores:
Quality: 104.50 Length: 273
Percent Similarity: 1.025 Gaps: 14
Percent Identity: 37.363

alignment_block:

US-09-303-518D-571 x YT44_STRFR ..

Align seg 1/1 to: YT44_STRFR from: 1 to: 395

```

156 AGACCGCGCGCGCATGTCGCCAATATGCG.....CGACGCGGCTT 196
32 ArgGlnArgGlnGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 48
197 TGAACCCGACACGACGAGCGTCGCAAGCGCTTTTGGGGAAGCGCAAA 246
48 lGlnAlaHisHisAlaAspGlnGlnGlnGlnGlnGlnGlnGlnGln 64
247 TCGGCTTGGAACCTGCCCCCGCTTTTCAAAAAACCGGAACATCGA 296
65 .....GluProValAspGlnHisGly 71
297 AACAAATGTCGCAAGCGGTGACAGCGCTGGAACACGTGCA..... 335
72 ProProCysLeuValGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 88
336 GCAGGCTTTGGCAAGCGGCGGCGCTGCTGTCATCAC..... 374
88 nArgGlyAspArgArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 405
375 .....GCCGACAT..... 383
105 lYArgProHisProPheGlnAlaLeuHisPheArgArgAspLeuPro 121
383 ..... 383
122 GlyValGlyValAspLeuLeuAlaArgGlyValGlnHisAspLeuGln 138
384 .....CGGACGTCGATTTGGCGGCGGACGCTACAT..... 413
138 gGlyArgGlnLeuArg...ArgArgProLeuHisAlaArgGlnArgLeu 154
414 .....CAGCCAGACGCTTCGCTCCACT 437
154 lLeuLeuLeuArgAlaLeuAlaLeuAlaLeuArgProGlyGlnGlnArg 170
438 GACCGCATGTACAAAGCGCGCA.....AATCAAGCGATAGACA 478
171 AspGlnHisLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 509
479 AATCATGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 509
187 uValHisGlyGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 204
510 .....AGGCAAAACCGCGCCACCGGCGATACAGGCGGTCAACAAATCAT 554
204 roAlaGlnGlnArgArgLeuAlaValHis..... 213
555 CAAGCGCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 604
214 .....ProValProGlyGln...LeuArgGlnProAlaArgArgTyr 227
605 CTTCCTCCGACGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 654
227 C.....ArgHisIleAsnGly..... 232
655 GCATACACCATGACACGCGGCGGCAAAATTTGGACACGTCGCAAGCGTGA 704
233 ..AlaHisSerGlnArgGlyAsp..... 239

```


462 rThrtPrtSertPValAlaGlyAspClyPthChlunArgLeuClYArgp 479

799 CACGA 803

479 roArg 480

seq_name: SwissProt_40:SON_MOUSE

seq_documentation_block: PRT; 2404 AA.
 ID SON_MOUSE STANDARD: PRT; 2404 AA.
 AC 09GXA7: 09GXP5: 09GCK6: 09CQ12;
 DT 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, last sequence update)
 DT 01-MAR-2002 (Rel. 41, last annotation update)
 DE SON protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC STRAIN=129/SV;
 RX MEDLINE=20408886; PubMed=10950926;
 RA Wynn S.L., Fisher R.A., Pagel C., Price M., Liu Q.Y., Khan I.M.,
 RA Zammit P., Dadgar K., Mazrani W., Kessling A., Lee J.S., Bulwela L.,
 RT "Organization and conservation of the GART/SON/DONSON locus in mouse
 RL and human genomes.";
 RL genomics 68:57-62(2000).
 RN [2]
 RP SEQUENCE OF 1-116 FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Hippocampus, Small intestine, and Tongue;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shigaawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schiml L.M., Stambli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Botfield D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamija M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mombauts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
 RA Wyshew-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 CC -1- FUNCTION: Transcriptional repressor. Binds to the consensus DNA
 CC sequence: 5'-GA[GC]A[NCG][AG]CC-3'. Might protect cells from
 CC apoptosis. Might be involved in pre-mRNA splicing (By similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are
 CC produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Widely expressed.
 CC -1- DOMAIN: Contains 8 types of repeats which are distributed in 3
 CC regions.
 CC -1- SIMILARITY: CONTAINS 1 G-PATCH DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 DBM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: AF193606; AAF23120.1;
 DR EMBL: AF193595; AAF23120.1; JOINED.
 DR EMBL: AF193596; AAF23120.1; JOINED.
 DR EMBL: AF193597; AAF23120.1; JOINED.
 DR EMBL: AF193598; AAF23120.1; JOINED.
 DR EMBL: AF193599; AAF23120.1; JOINED.
 DR EMBL: AF193600; AAF23120.1; JOINED.
 DR EMBL: AF193601; AAF23120.1; JOINED.
 DR EMBL: AF193602; AAF23120.1; JOINED.
 DR EMBL: AF193603; AAF23120.1; JOINED.
 DR EMBL: AF193604; AAF23120.1; JOINED.
 DR EMBL: AF193605; AAF23120.1; JOINED.
 DR EMBL: AF193607; AAF23121.1;
 DR EMBL: AK019312; BAB31659.1;
 DR EMBL: AK019081; BAB31536.1;
 DR EMBL: AK008478; BAB25691.1;
 DR EMBL: AK008256; BAB2562.1;
 DR MGI: 96353; Son.
 DR InterPro: IPR001159; DS_RBD.
 DR InterPro: IPR000467; G_patch.
 DR Pfam: PF00035; dsrm; 1.
 DR Pfam: PF01585; G_patch; 1.
 DR SMART: SM00443; G_patch; 1.
 DR PROSITE: PS50137; DS_RBD; 1.
 DR PROSITE: PS50174; G_PATCH; 1.
 KW RNA-binding; DNA-binding; Nuclear protein; Repeat;
 KW Alternative splicing.
 FT EMBL 721 850
 FT EMBL 867 943
 FT EMBL 961 1080
 FT EMBL 961 966
 FT EMBL 969 974
 FT EMBL 976 981
 FT EMBL 985 990
 FT EMBL 993 998
 FT EMBL 1001 1006
 FT EMBL 1010 1015
 FT EMBL 1018 1023
 FT EMBL 1026 1031
 FT EMBL 1035 1040
 FT EMBL 1044 1049
 FT EMBL 1055 1060
 FT EMBL 1066 1071
 FT EMBL 1075 1080
 FT EMBL 1101 1133
 FT EMBL 1910 1979
 FT EMBL 1910 1916
 FT EMBL 1938 1944
 FT EMBL 1945 1951
 FT EMBL 1952 1958
 FT EMBL 1959 1965
 FT EMBL 1966 1972
 FT EMBL 1973 1979
 FT EMBL 1919 1990
 FT EMBL 1919 1937
 FT EMBL 1980 1990
 FT EMBL 1991 2017
 FT EMBL 2283 2329
 FT EMBL 2349 2404
 FT EMBL 2086 2086
 FT EMBL 2087 2404
 SO SEQUENCE 2404 AA; 261428 MW; 64BFB28BD3FC0D9 CRC64;

alignment_scores: Quality: 103.50
 Ratio: 1.035

Length: 201
 Gaps: 7

Percent Similarity: 49.751 Percent Identity: 25.373

alignment_block:

US-09-303-518d-571 x SON_MOUSE

Align seg 1/1 to: SON_MOUSE from: 1 to: 2404

```

230 TTGCGGAACGCAAAATCGGTTTGGAACCTGGCCCCGCTTTTCAAA 279
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1807 LeuArgSerArgSerArgSerArgSerArgSerArgSerArgSerArg 1823
280 AAACCGGAGACATCGAACAATGTTCAAGCGGTACAGCGCTGGAGACA 329
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1823 sArgThrSerGluSerArgSerArgAlaArgLysArgSerSerL 1840
330 CGTGACGAGCGTTTGACAGAGCGGCGTCTGTTCAACGCGCGC 379
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1840 ySerHisArgSerGlnThrArgSerArgSerArgSerArgArgArg 1856
380 ACATCGGACAGCTACGATTGGCGGACGCTACATCAGCCAGCGCTTCG 429
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1857 ArgSer.....SerArgSerArgSerly 1864
430 TTCACCTGACCGCCGATGACAGCCCGCAAAATCAAGCATACACA 479
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1864 sSerArgGlyArgArgSerArgSerArgSerArgSerArgSerArg 1881
480 AATCATCAGGCGGCGGCGGTGCGCGCAAAAGCAACCGCGCCACCG 529
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1881 yS.HisArgSerLysSerArgGluArgLysArgLysArgSerSerArg 1897
530 GCATACAGG.....GTCACAAACAATCATCAGCGCTGCGCGCGCGC 573
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1897 gAspAsnArgLysAlaAlaArgAlaArgSerArgSerArgSerArg 1914
574 GAGCAGCAACATCATCTGCC...CGACACAGCTCTCTCCGAGGAGAG 620
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1914 eArgSerHisThrProSerArgArgArgArgSerLysSerValGlyArg 1930
621 CGGCGCGCTGTGGCGGATTTTTCGCAAACTGCTACATCAGCATGAC 670
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1931 ArgArgSerPhe..... 1934
671 TGGCGGCAAAATGTCACAGCTCAAGCGCTGTAACCTGTTTCTGCG 720
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1949 eArgThrProSerArgArgSerArgThrProSerArgArgSerArgThr 1965
721 TGGCAGCGCTGCGCGACGAGCAAGCCTT.....CGTGTTGCACAT 761
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1949 eArgThrProSerArgArgSerArgThrProSerArgArgSerArgThr 1965
762 CCGCCCCCTCCCAAGGGAATTGAAGCGGACCAAGCCAGATGCCCGC 811
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1966 ProSerArg.....ArgSerArgThrProSerArgArgArg 1977
812 T 812
1977 g 1977
seq_name: SwissProt_40:SFR2_CHICK
seq_documentation_block:
ID SFR2_CHICK STANDARD: PRT; 221 AA.
AC P30352:
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Splicing factor, arginine/serine-rich 2 (Splicing factor SC35) (SC-35)
DE (Splicing component, 35 kDa) (PR264 protein).
GN SFRS2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

```

```

OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thymus;
RX MEDLINE=92212859; PubMed=1557353;
RA Vellard M., Sureau A., Soret J., Martinierie C., Perbal B.;
RT "A potential splicing factor is encoded by the opposite strand of the
RL trans-spliced c-myc exon.";
RC Proc. Natl. Acad. Sci. U.S.A. 89:2511-2515(1992).
CC -1- FUNCTION: NECESSARY FOR THE SPLICING OF PRE-MRNA. IT IS REQUIRED
CC FOR FORMATION OF THE EARLIEST ATP-DEPENDENT SPLICING COMPLEX AND
CC INTERACTS WITH SPLICOSOMAL COMPONENTS BOUND TO BOTH THE 5' AND 3'
CC SPLICING SITES DURING SPLICOSOME ASSEMBLY. IT ALSO IS REQUIRED FOR
CC ATP-DEPENDENT INTERACTIONS OF BOTH U1 AND U2 SNRNPs WITH PRE-
CC mRNA.
CC -1- SUBCELLULAR LOCATION: Nucleus.
CC -1- ALTERNATIVE PRODUCTS: DIFFERENT FORMS OF THE PROTEIN MAY BE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE RS
CC DOMAIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X62446; CAA4306.1; .
DR PIR: S17327; S17327.
DR PIR: B42701; B42701.
DR HSSP: P09651; 1UP1.
DR InterPro: IPR00504; RRM.
DR Pfam: PF00076; rrm; 1.
DR SMART: SM00360; RRM; 1.
DR PROSITE: PS50102; RRM; 1.
DR PROSITE: PS00030; RRM_RNP_1; 1.
DR Nuclear protein; RNA-binding; mRNA splicing; Alternative splicing;
KW Phosphorylation.
FT DOMAIN 14 92 RNA-BINDING (RRM).
FT DOMAIN 111 116 GLY-RICH (HINE REGION).
FT DOMAIN 117 221 ARG/SER-RICH (RS DOMAIN).
FT SEQUENCE 221 AA; 25524 MW; 75A4DBF9170F1BF CRC64;
SQ

```

alignment_scores: Quality: 102.50 Length: 211
Ratio: 0.995 Gaps: 8
Percent Similarity: 48.815 Percent Identity: 22.749

alignment_block:

US-09-303-518d-571 x SFR2_CHICK

Align seg 1/1 to: SFR2_CHICK from: 1 to: 221

```

238 ACTTGGCCCCGCTTTTCAAAAACGGAAGACATCGAACAATGTTCA 307
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
29 ThleuArgArgValPheGluLysTyrGlyArgValGlyLysPValTyr 45
308 AAGCGGTACAGCGCTGGGAGACAGTGCAGCGCTTTGA..... 347
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
45 eProArgAspArgTyrThrLysGluSerArgGlyPheAlaPheValArgP 62
347 ..... 347
62 heHisAspLysArgAspAlaGluAspAlaMetAspAlaMetAspGlyAla 78
348 .....CAAGGCGAAGGCGCTGTTTCATCATCAGCGCGCACATCGGACGCT 391
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

79 ValLeuaspGlyArgGluLeuArgValGlnMetala..... 90
392 AGCATTTGGGCGAGCTACATCACCAGCAAGCTTCCCTCCAGTACC 441
   |||||
91 ArgTyrGlyArg..... ProProsp 98
442 GCCATGTACAGCCCGCAAAATCAGATAGCAAAATCAGCAAGC 491
   |||||
98 erHis...HiserArgGlyProProProAlaArgTyrGlySer 113
492 GGGCAGGGGCGCGCAAGCAAGCAAGCCCGCCAGCATACAGAGG 541
   |||||
114 GlyTyrGlyArgSerArgSerArgSerArgArgArgArgSerArg 130
542 TCACCAATCATCAGAGCCCTCGCGCGCGCGAGCAACATCAGCTG 591
   |||||
130 rArgSerArgSerArgSerArgSerArgTyrSerArg 147
592 CCCGACGACGTCCTTCCTCCGAGAGAGCGCGCGGATTT 641
   |||||
147 erLysSerArgSerArgTyrArgSerArgSerArg..... 158
642 TTTGGCAACCTGCATACACATGACACTGCGCAAAATTCGACAG 691
   |||||
159 .....SerThrSerLysSerArgSerAla 167
692 TCACAGCGTGAACCCCTTTTCTGCTGCGAGCAGCTCCGCA...C 738
   |||||
167 rArgSerLysSerLysSerSerValSerArgSerArgSerArg 184
739 GACACAGCTTCTGTGACATCCGCGCGCGCAAGGGGATTCAGCG 788
   |||||
184 rArgSerArgSerArgSerArgSerPro...ProThrSer...LysArg 198
789 CACCAAGCCGACGATCCGCGCTGTTCAGCG 821
   |||||
199 GluSerArgSerArgSerArgSerPro 209

seq_name: SwissProt_40:FXR2_HUMAN

seq_documentation_block:
ID FXR2_HUMAN STANDARD; PRT; 673 AA.
AC P51116;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fragile X mental retardation syndrome related protein 2.
GN FXR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96080171; PubMed=7489725;
RA Zhang Y., O'Connor J.P., Sloni M.C., Srinivasan S., Dutra A.,
RA Nusbaum R.L., Dreyfuss G.;
RT "The fragile X mental retardation syndrome protein interacts with
RT novel homologs FXR1 and FXR2."
RL EMBO J. 14:5358-5366(1995).
RN [2]
RP SEQUENCE OF 1-27 FROM N.A.
RX MEDLINE=98100442; PubMed=9437788;
RA Joseph D.R.;
RT "The rat androgen-binding protein (ABP/SHBG) gene contains triplet
RT repeats similar to unstable triplets; evidence that the ABP/SHBG and
RT the fragile X-related 2 genes overlap."
RL Steroids 63:2-4(1998).
CC -1- FUNCTION: RNA-BINDING PROTEIN. INTERACTS WITH FXR1 AND FXR2.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: CONTAINS 2 KH DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE FXR1 FAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DR EMBL: U13501; AAC50292.1;
DR EMBL: AF044263; AAC03357.1;
DR HSPB: Q06787; 2FMR.
DR MIM: 605339;
DR InterPro: IPR004087; KH.
DR InterPro: IPR004088; KH_TYPE_1.
DR Pfam: PF00013; KH-domain; 1.
DR SMART: SM00322; KH; 2.
DR PROSITE: PS5084; KH_TYPE_1; 2.
DR RNA-binding; Repeat.
FT DOMAIN 232 261 KH 1.
FT DOMAIN 295 324 KH 2.
FT DOMAIN 414 418 POLY-SER.
FT DOMAIN 544 552 POLY-ARG.
FT DOMAIN 584 594 POLY-ARG.
SQ SEQUENCE 673 AA; 74128 MW; D20FBD8634D0B CRC64;

alignment_scores:
Quality: 102.50 Length: 223
Ratio: 0.967 Gaps: 10
Percent Similarity: 47.534 Percent Identity: 23.767

alignment_block:
US-09-303-518d-571 x FXR2_HUMAN
Align seg 1/1 to: FXR2_HUMAN from: 1 to: 673

192 GGGTTGAACCCCGACGACGACGCTCAAAAGCCGTTTTCGGAAGACG 241
   |||||
391 GlyPheArgProPro.....GlySerG 398
242 CAAATGCGTTGGAACTGCCCCGCGTTTCAAAAACCGAGAGAC 291
   |||||
398 rArgGlySerGlySer.....AspLysAlaLys 410
292 ATCGAAACAAATGTTAAAGCCGTACAGCGCTGGAGACGTCGACAGC 341
   |||||
410 erThrAspGluSerSerSerSerSerLeuHisAlaThrArgThrGly 426
342 TTTGACAGAGGCGAGAGGCGCTGTTCATCAGCCCGCACATGGCAGCT 391
   |||||
427 GlySerTyrGlyArgGly.....ArgLys 436
392 AGCATTTGGCGCGAGCTACATCAGCCAGCGCTTCGCTCCAGCTGACC 441
   |||||
436 rArgThrGlyGlyProAlaTyrGlyProSerSerArgValSer..... 450
442 GCCATGTACAGCCCGCAAAATCAGATAGCAAAATCAGCAAGC 491
   |||||
451 .....ThrAlaSerGluThrLysSerLys..... 459
492 GGGCAGGGTCCGCGCAAGCAAGCAAGCCGCGC.....CA 526
   |||||
460 .....ArgGluGluProAsnArgAlaGlyProGlyAspArgAs 472
527 CCGCATACAGAGGAGCAAAATCATCAGAGCCCTCGCGCGGAGAG 576
   |||||
472 ProProThrArgGlyGluGluSerArgArgArgProThrGlyArgG 489
577 GCACATCATCTCTGCGCGACGAGCTCTCCGAGAGAGGCGGAG 626
   |||||
489 LysArgGlyProProAlaProAlaProThrSerArgTyrAsnSer 505

```

627 CGTGGGCGGATTTTTCGCAACCTGCATACACGACACTGGCG 676
 506 SerIleSerValLeuLysAspPro..... 514
 677 CAAATTTGGCAGCTCAAGGCGTGAACCCGTTTCTGCTGCA 726
 515AspSerAsnProTyrSerLeuLeuAsp 524
 727 CG.....CTGCGCGAGCAGAGCGTGTGTCATCGCCCGCT 770
 524 hSeSerGluProGluProValAspSerGluProGluProPro... 539
 771 CCAAGGCGATTTGAAGCGCAACAAACCCGCGCTGTTCAAC 820
 540 ProAlaSerAlaArgArgArgSerArgArg...ArgArgThrAsp 555
 821 GCATATCGCATATTTGAT 839
 555 uAspArgThrValMetAsp 561
 seq.name: SwissProt_40:VE2_HPV36

seq_documentation_block:
 ID VE2_HPV36 STANDARD; PRT; 509 AA.
 AC P50809;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Regulatory protein E2.
 GN E2.
 OS Human papillomavirus type 36.
 OC Viruses; dsDNA viruses, no RNA stage: Papillomaviridae;
 OC Papillomavirus.
 OX NCBI_TaxID=37957;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA deLius H.;
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
 CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACNNNNNGGT-3') PRESENT
 CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
 CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDENT OF E2RE'S POSITION
 CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
 CC BY SPERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
 CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
 CC REPLICATION.
 CC -1- SUBUNIT: BINDS DNA AS A DIMER.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: U31785; AAA79439.1; -
 DR HSSP: P03122; 280P
 DR InterPro: IPR001866; E2_N
 DR InterPro: IPR000427; Early2_C
 DR Pfam: PF00511; E2_C; 1.
 DR Pfam: PF00508; E2_N; 1.
 DR ProDom: PD000672; Early2_C; 1.
 DR ProDom: PD000678; E2_N; 1.
 DR Early protein: Transcription regulation; Activator; DNA-binding;
 KM Trans-acting factor; DNA replication; Repressor; Nuclear protein.
 SQ SEQUENCE 509 AA; 57186 MW; BD41FA299E150D91 CRC64;

alignment_scores:
 Quality: 102.00
 Ratio: 0.836

Length: 241
 Gaps: 10

Percent Similarity: 50.622 Percent Identity: 22.822

Alignment_block:
 US-09-303-518D-571 x VE2_HPV36

Align seg 1/1 to: VE2_HPV36 from: 1 to: 509

185 GGCAGGCGGTTTGAACCCGCGACGCGGTCAACGCGTTTGGC 234
 250 GYAArgArg.....ProSerArgThrArgArgProAlaLal 263
 225 GAACGCGCAAAATCGGTTTGAACCTTGCCCGCTTTTCAAAAAC 284
 263 SGLnArgArgSerArg.....SerArgHisArgSerSer 275
 285 GGAACATCGCAAAATCGTTTCAAGCGGTACACG..... 319
 275 rSerArgSerArgSerArgSerProSerLeuAlaLysThrGluAla 291
 320GCTGGACACGTCGACGAG 339
 292 AlaThrThrArgSerArgSerProSerLeuAlaLysThrGluAla 308
 340 GCTTTGGACAGGCGGAGGCTGCTGTTATCAAGCGCGCATCGGACG 389
 308 gValSerThrArgSerArgSerArgSerThrSerArgArgGlyArg 325
 390 CTACGATTTGGCGGACGCTACATCAGCAGCAGCTTCCGTTCCACTGA 439
 325 rArgArgSerArgSerProSerThrSerSerThrThrThrAs 341
 440 CCGCATGTACAAACCCGCAAAATCAAGCATGACAAATCATCGAG 489
 341 nLysArgSerArgValArgAlaLysThrThrGlySerArgGlyAlaArg 358
 490 GCGGCGAGGCGCGCGCAAGCAAAACCGCGCCACCGCATACAAAG 539
 358 lGlyArgGlyAlaArgGlyGlySerGlyGlyArgArgArgGlyArg 374
 540 GGTCAACAATCATCAAGCGCTGCGCGCGCGGCGGACACATCATC 589
 375 SerSerSerSer...ThrSerProAlaHisLysArgSerArgGluHis 390
 590 TGCCGCGACGCTCCCTTCCGACAGAAAGCGCGCGCTGGCGGAT 639
 390 rValArgSerArgGlyValSerProAsp.....GluValGlyLys 404
 640 TTTTGGCAACCTGCATACCATGACACCTGCGCAAAATTGCG... 686
 404 rLeuAlaArgSerValSerSerLysHisThr.....GlyArgLeuGlyArg 418
 687 ...ACAGCTCAAAAGCGGTGAACCCGTTTTCGCTCGCAACGCTCG 733
 419 LeuLeuGluGluAlaLeuAspProValIleLeuValArgGlyGlu 435
 734 CGGA.....CGGACAAAGCGTTCGTTGTCATCCGCCCC 768
 435 aAsnThrLeuLysCysPheArgAsnArgAlaLysIleLysTyrMetGly 452
 769 GTCGCAAG.....GGAATT 782
 452 eUTyArgSerPheSerThrThrTrpSerTrpValAlaGlyAspGlyThr 468
 783 GAACGCAACAAAGCCACGA 803
 469 GluArgLeuGlyArgProArg 475
 seq.name: SwissProt_40:YH1_EBV
 seq_documentation_block:
 ID YH1_EBV STANDARD; PRT; 660 AA.
 AC P03181;
 DT 21-JUL-1986 (Rel. 01, Created)

21-JUL-1986 (rel. 01, Last sequence update)
 15-JUL-1998 (rel. 36, Last annotation update)
 DE Hypothetical BHLPI protein.
 OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 NCBI_TaxID=10377;
 RN [1]

SEQUENCE FROM N.A.
 MEDLINE=84270667; PubMed=6087149;
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
 RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
 RA Tuffnell P.S., Barrell B.G.,
 RL "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
 Nature 310:207-211(1984).

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL, V01555; NOT_ANNOTATED_CDS.
 DR PIR, A03742; OQBE3.
 KM Hypothetical protein: Early protein; Repeat.
 FT DOMAIN 148 648 4 X 125 AA TANDEM REPEATS.
 FT REPEAT 149 273 1.
 FT REPEAT 274 398 2.
 FT REPEAT 399 523 3.
 FT REPEAT 524 648 4.
 SO SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;

Alignment_scores:

Quality: 102.00 Length: 306
 Ratio: 0.779 Gaps: 13
 Percent Similarity: 42.810 Percent Identity: 25.163

alignment_block:

US-09-303-518D-571 x YHLL_EBV

Align seg 1/1 to: YHLL_EBV from: 1 to: 660

```

27 TCCCTCTTGGCAAGCCGACATCTGTGACCGCCGCT..... 71
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
211 ThrProHisProGluArgGlySerGlyProAlaAspProGluAlaAla 227
72 .....CAAATGCTCTCCCTGCTGCG 93
227 aArgLeuProProGluArgGlnGluProArgLeuProGlnAspLeuAla 244
94 GTTTCCTGTGTCACAGCTGGGAACCGGCTCGACATCTGGGCTTTA 143
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
244 lAlaGlnArgGlyProAlaGlyProProProThrArgSerGlyAla... 259
144 CCTTTAAGAGAGACCGCGCGCATGTGCGCAATATGCGAGCGCG 193
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
260 .....AlaAlaGlnArgThrHisArgArgProGlyGly... 271
194 GTTTAACCCGACAGCAGCGGTCAAAGCGCTTTTGGGGAACGCGCA 243
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
272 .....ProArgSerAla..... 275
244 AATGCGGTTTGAACCTTGGCCCGCTTTTCAAAAACCGGAGACAT 293
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
276 .....ArgAsnProGlyCysProArg.....ThrTrpArgGly 286
294 CGAAACATGTTCAA..... 308
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
287 ArgSerGlyAlaGlnArgGlnHisProProProGlyAlaGlnArgPr 303

```

seq_name: SwissProt_40:SGS4_DROME

seq_documentation_block:

SGS4_DROME STANDARD; PRT; 297 AA.

AC 000725; OSN472;
 DT 01-OCT-1996 (rel. 34, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Salivary glue protein Sgs-4 precursor.
 GN Sgs4 OR SGS-4 OR SG:96G10.6 OR CG12181.
 OS Drosophila melanogaster (Fruit fly)
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]

SEQUENCE FROM N.A.

RP STRAIN=OREGON-R, KARSNAS, AND SAMARKAND;
 RX MEDLINE=92223113; PubMed=1562607;
 RA Furia M., Digilio F.A., Ariano D., Favia G., Polito L.C.;
 RT "Molecular characterization of a Drosophila melanogaster variant
 RT strain defective in the Sgs-4 gene dosage compensation.";
 RL Biochim. Biophys. Acta 1130:314-316(1992).
 RN [2]

RP SEQUENCE FROM N.A.

```

309 .ACCGTACACGGCTGGGAACACGTGCACAGCGCTTTGGCAAGCGCGAA 357
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
303 oSerGlyProThrGlyGlyArgProAlaAlaProGlyAlaProGlyThrP 320
358 GGGCTCTGTTCATACCGCCGACATCGGACGCTTCGATTTGGGCGGACG 407
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
320 roAlaAlaProGlyProGlyGlyGlyAlaAlaValProSerGlyAlaThr 336
408 CTACATCACCACGACAGCTTCGTTCCACTTCACCGCATGTACAGCC.. 455
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
337 ProHisProGluArgGlySerGlyProAlaAspProProAlaAlaAla 353
456 .....GCCGAAATCAAGCATGAG..CAAATCATCGACGCG 492
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
353 gluProProGluArgGlnGluProArgLeuProGlnAspLeuAlaAla 370
493 GGCAGGG..TGGCGGCAAAAGCAAAACCGCCCGCCGCGCATCAAGGG 541
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
370 lArgArgCysProAlaGlyProProProThrArgSerGlyAlaAlaAla 386
542 TCMAACATCATCAGAGCC.....TGGCGCGGCGGCGAGCAACCATC 585
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
387 GlnArgThrHisArgArgProProGlyCysProArgSerAlaArgAsnPr 403
586 ATCCGCGCCGACACGTCCTTCGCGCAGGAAGCGCGCGCTGGCG 635
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
403 oGlyCysProArgThr...TrpArgArgSerGlyAlaGlnArgGly. 418
636 GGAATTTTTCGCAAACTGCTATACACCATGACACGTGGCGCAAAATTTG 685
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
419 .....HisProProGluArgGlyAlaGlyGlnArgPro 428
686 CACACGTCAAAAGCGGTGAAAACCTGTTTTCGCGCAGCAAGCGCGCC 735
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
429 SerGlyProThrGlyArgProAlaAlaProGlyAlaProGlyThrP 445
736 GACGACAAAGCT.....TCGTGTCACAT 761
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
445 oAlaAlaProGlyProGlyGlyAlaAlaAlaValProSerGlyAlaThrP 462
762 CGCGCCGTCGCAAGGGAATTGACGCAACAGCCGACGATGCGCGCG 811
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
462 roHisProGluArgGly.....SerGlyProAlaAspProPro 474
812 TGTTCACCGCAATACCG 829
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
475 AlaAlaAlaArgLeuPro 480

```



```

151 |:::|:::|:::|
|gthrghlupProthrcysLysThrghlupProthrcysLysThrghlup 168
264 |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
168 |roProthrcysLysThrghlupProProCysGluYshLysThrghlup 184
314 |TACACGCGTGGGACACGTCGACGAGCTTGGACAGGGGAGG... 358
185 |IleLysArgHisArgThr.....LysArgThrLysArgSerLysSerTh 199
359 |.....G 359
199 |LysLysIleValHisHisHisAsnArgProGlyThrThrProGlySerG 216
360 |GCTGC...TGTTCATCAGCGCCGACATCG.....GAGCTACG 394
216 |LysGlyCysGlySerLysAsnGluSerLysLysLysSerGlyCys 232
395 |ATTGGGCGGACGCTACATCAGCCAGCAGCTCCGTTCCACCTGACCGCC 444
233 |IleLysAspLeuLeuThr.....Proly 241
445 |ATGTACAGCGCGCGGAAATCAAGCGATGACAAATCATGACG... 490
241 |CysProAspSerLysProLysProGlnAlaSerProLysCysLysSerA 258
491 |.....CGGCGAGGCTGCGCGGCAAGCAAGCAAGCGCGCACCG 529
258 |sProlysProLysAlaAlaSerLysThrSerLysProLysProLys 274
530 |GCATACAGGCGGTCAACAATCATCAAGCGCCCTGCGCGGCGACGCA 579
275 |AlacysAspSerClyLysLysAsnThrThrLysLysProArgLysThrG 291
580 |ACCA 583
291 |nPro 292

```

seq.name: SwissProt_40:MANA_RHOMR

seq_documentation_block:

```

ID MANA_RHOMR STANDARD; PRT; 1021 AA.
AC P49425.
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mannan endo-1,4-beta-mannosidase (EC 3.2.1.78).
GN MANA.
OS Rhodothermus marinus (Rhodothermus obamensis).
OC Bacteria; CFM group; Rhodothermus group; Rhodothermus.
OX NCBI_TaxID=29549;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 43812;
RA Pollitz O., Krah M., Bortless R.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic
CC linkages in mannans, galactomannans, glucomannans, and
CC galactoglucomannans.
CC -1- SIMILARITY: BELONGS TO FAMILY 26 OF GLYCOSYL HYDROLASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
DR EMBL; X90947; CA62442.1; -
DR InterPro: IPR000805; Glyco_hydro_26.
DR Pfam; PF02156; Glyco_hydro_26; 1.

```

DR PRINTS: PR00739; GLHYDRLASE26.
KW Hydrolase; Glycosidase.
SQ SEQUENCE 1021 AA; 115791 MW; A8F7B63109E0F715 CRC64;

alignment_scores:
Quality: 101.50 Length: 268
Ratio: 0.839 Gaps: 15
Percent Similarity: 45.149 Percent Identity: 24.627

alignment_block:

US-09-303-518D-571 x MANA_RHOMR ..

Align seg 1/1 to: MANA_RHOMR from: 1 to: 1021

```

228 |TTTGGCGAAGCGCAAAATGCGGTTTGGAACTTGCCCGCGCTTTTGA 277
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
33 |PheAlaGlyHisGlyLeu.....AlaAspGlnArgLeuIleAl 46
278 |AAAGCGGAAAGACATCGAAACATGTTCAAGCGGTACAGCGCTGGAA 327
46 |HisThrGlyArgProAlaLeuAlaGlnAlaGlnArgIleThrGlnGlnLysG 63
328 |CAGCTGACGAGCGCTTGGACAAAGCGCGAAGCGCTGCTTGCATCAGCG 377
63 |LysArgSerGlnCysArgGlyGlnArgHisArg.....HisGln 76
378 |GCACATCGCGAGCTACGATTTGG...CGAGCGTACATCAGCGACG 424
77 |GlyGlyArgAsnLeuArgLysAlaHisArgThrPheHisGlnHisGlnSe 93
425 |TTCCGTTCCACCTGACCGCGCATGTACAAAGCGCGCAAAATCAGCAT 474
93 |Tyr.....GlnAspGlnAlaHisAspAlaProHisGlyGlnGlnAla 108
475 |GACAAATCATGACGAGC...GGCGAGGCGCGCGCAAGCAAGCGCG 521
108 |YThrGlyHisGlnGlyLeuGlyLeuGlyHisGlnHisAlaGlnAlaGln 122
522 |GCCACCGCATCAAGGGGTCAACAAT..... 551
123 |.....HisGlnGlnGlyGlnSerAsnValAlaAspArgGlnAspG 136
552 |.....CATCAAGCGCGCTGCGCGCGCGGAGCGCAACC 582
136 |YglupProValGlnAlaGlnHisGlnLysAspGlyAlaGlnArgAlaGly 153
583 |ATCATCGTCCG.....CGACCACTCCCTTCTCCGCA... 614
153 |SnAlaProAlaGlyArgValGluLeuGlnGlnGlnProValGlnAlaGln 169
615 |.....GGAAGCGCGCGCGTGTGG 634
170 |HisGlnGlnGlnGlnGlnGlnAlaArgIleGlyLysArgArgGln..... 184
635 |CGAGTTTTCGCGAAACCTGCATACACCATGACACAGCGCGCAAAATG 684
185 |AspAlaPheAlaProProAlaLeuAspHisValHisGlyPro.... 199
685 |GCACACGTCGAAGCGTGAAGACCTGTGTTTCTGCTGCGACGCTGCG 734
200 |GlyArgLeuGlnAlaGlnHisGlyLeuAlaValGlnArgHisValProAla 215
735 |.....CGACGACAAAGCGTTCGTTGTCACAA 760
216 |ValGlnGlnHisGlnGlnAlaGlnAlaGlnArgGlyArgGlnGlnIleAsp 232
761 |..... 761
232 |SValLeuGlyHisGlyLeuProGlyArgGlnArgLeuAlaPheArgAsp 249
762 |..CGCGCGCTCCAGCGGAATGAAACGCAACAAAGCGCGACAGTGC... 806
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

1

RP SEQUENCE FROM N.A.
RX MEDLINE=80058903; PubMed=3680218;
RA Rothnagel J.A., Mehrel T., Ridler W.W., Koop D.R., Steinert P.M.;
RT "The gene for mouse epidermal filaggrin precursor. Its partial
RT characterization, expression, and sequence of a repeating filaggrin
RN unit.";
J. Biol. Chem. 262:15643-15648(1987).
[2]

RA REVISIONS. J. A.:
 RP Rothnagel J. A.:
 RL Submitted (SEP-1988) to the EMBL/GenBank/DBJ databases.
 CC -I- FUNCTION: AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES
 CC DISULFIDE-BOND FORMATION AMONG THE INTERMEDIATE FILAMENTS DURING
 CC TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.
 CC -I- PPM: FILAGGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE,
 CC HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM COPIES
 CC OF 248 AA, WHICH ARE NOT SEPARATED BY "LARGE LINKER". THE
 CC PRECURSOR IS DEPOSITED AS KERATOHYALIN GRANULES. DURING TERMINAL
 CC DIFFERENTIATION IT IS DEPHOSPHORYLATED & PROTEOLYTICALLY CLEAVED.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J03458; AAA75559.1; -
 DR PIR: A28444; A28444.
 DR MGD: MGT:95553; Flg.
 KM Phosphorylation; Developmental protein.
 FT NON_TER 1
 FT 1
 SQ SEQUENCE 336 AA: 35678 MW: 259F124D3AC0B2D CRR6A.

```
alignment_scores:
  Quality: 100.50
  Ratio: 0.756
Percent Similarity: 46.503
Percent Identity: 21.329
```

```
alignment_block:
```

Align seg 1/1 to: FILA_MOUSE from: 1 to: 336

[illegible]

```

362 TCCCTTATCATCCGSCCA.....CATGGCGCTTCAGT 396
397 TTGGCGGAGCGTTACATCAACGACGCGTCCGTTCCAGTACGCGGAT 446
105 YALSERGLSERARGLASERASPSERGLYHISERASPSEREG 122
122 LUGLYGALALVALGLYALHISARGLINSERGLYALGLYGLNARGHIS 138
447 GPACAGGC..... 455
139 GLUGLINSERSEARGLYGLNHSGLYSERGLYTYTYTYGLUCL 155
456 .....GCCAAATCAAGCGATAGACAAATCATGACGCGGCGAGG 498
155 NGLHISERGLUGLINSERASPSERGLNHSGLNHSGLYHISGLNH 172
499 GTGGCGGCAAGG.....CAAAACGCGCGCCACGCGCAT 533
172 LSGLUGLINSARGLYHISGLNHSGLNHSGLNHSGLNHSGLNHS 188
534 ACAAGCGGTCAAACATCAT...CAAGCGCGTCCGCGCGCGAGCA 580
189 GLULNRPGILUSERGLYHISARGLINSERGLYARGLYH 205
581 C.....CATCATCTGCCCGGACACGCTGCC 606
205 SGLGLYALHISGLNGLYGLNGLYARGASPSERALARGPROARGLYS 222
607 TCTCCGAGGAAGCGCGCGCTGGCGGATTTTGGCAACCTGC 656
222 ERASNGLGLYHISERSE.....ARGHISGLNLA 233
657 ATACACCATGACACTGCGCGCAAAATGGACACGTCAAAGCGCTGA 706
234 ASPSERPROARGVALSERALARGSERGLYSERGLYARGLYGLNSE 250
707 CCTGTGTTTTGCTGCGAACGCTCCCGACGAGCAACGCTGCTGTG 756
250 RPROASRALASERGLYARGSERASPSERASPSARGASPSARGPROARGLNP 267
757 CACATCG 764
267 ROSERPRO 269

seq_name: SwissProt_40:LIPD_AERHY

seq_documentation_block:
ID LIPD_AERHY STANDARD; PRT; 684 AA.
AC P40600;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Extracellular lipase precursor (EC 3.1.1.3) (Triacylglycerol lipase),
GN LIP.
OS Aeromonas hydrophila.
OC Bacteria, Proteobacteria; gamma subdivision; Aeromonadaceae;
OC Aeromonas.
OX NCBI_TaxID=644;
RX MEDLINE=93378423; PubMed=8368830;
RA Angula J., Rodriguez Aparicio L.B., Naharro G.;
RT "Purification, gene cloning, amino acid sequence analysis, and
RT expression of an extracellular lipase from an Aeromonas hydrophila
RT human isolate."
APPL. Environ. Microbiol. 59:2411-2417(1993).
RC STRAIN=H3;
RN SEQUENCE FROM N.A.
CC -1- FUNCTION: THE OPTIMUM CHAIN LENGTHS FOR THE ACYL MOIETY IS C6 FOR
CC ESTER HYDROLYSIS AND C6 AND C8 FOR TRIACYLGLYCEROL HYDROLYSIS. THE
CC OPTIMAL TEMPERATURE FOR ENZYME ACTIVITY IS 37 DEGREES CELSIUS.
CC HOWEVER, AFTER 30 MIN. AT 55 DEGREES CELSIUS, THE ENZYME REMAINS
CC 90% OF THE LIPASE ACTIVITY.

```

CC -1- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a
 CC fatty acid anion.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: S65123; AAB28083.2; -
 DR InterPro: IPR000734; Lipase.
 DR PROSITE: PS00120; LIPASE_SER; 1.
 KW Hydrolyase; Lipid degradation; Signal.
 FT SIGNAL 1 48 POTENTIAL.
 FT CHAIN 49 684 EXTRACELLULAR LIPASE.
 FT ACT SITE 568 568 CHARGE RELAY SYSTEM (BY SIMILARITY).
 SO SEQUENCE 684 AA; 71903 MW; 325418D3936FMA428 CRC64;

alignment_scores:
 Quality: 100.50 Length: 255
 Ratio: 0.966 Gaps: 14
 Percent Similarity: 40.784 Percent Identity: 24.706

alignment_block:

US-09-303-518D-571 x LIPE_AERHY

Align seg 1/1 to: LIPE_AERHY from: 1 to: 684

42 CGCCATGCACATCT...GTTGACCGCTCTCAAAATGCTCTCCT... 86
 272 ArgHisLeuYrProGlyArgAspArgProThrCysLeuProAspProAl 288
 87GCTGTGCG 93
 288 aglyCysArgAlaValProAlaIaGlInLysAspAlaIleAlaThra 305
 94 CTTTCCTGTGTGCACAGCGTGGAAACCGCTCGACATCTGCGCTTTTA 143
 305 Iaphe.....GlyProValLeuArgSerThrArgLeuLeu 316
 144 CTTTAAAGAGAACCGCGCGG..... 167
 317 Lys.....ArgProArgSerIleProValProSerSerCysLeuTh 330
 168CATCGTCGCAATATGCGGAGCGGCTTTGAAACCGCGAC 207
 330 rSerSerProHisArgArgProGlnValProGlyAla...ArgProArgP 346
 208 AGCGAGAGCGGTAAAGCGCTTTTGGGAAGCGAAATGCGGTGGA 257
 346 rSerPro.....Gly 349
 258 ACTGCGCCCGCGCTTTTCAAAAACCGAGAGACATCGAAATGTTCA 307
 350 ThrValPro.....SerGlnProValArgHisArgGlnCysAlaG1 363
 308 AAGCGGTACAGCGTGGAGAACAGCTGACAGCGCTTTGAGAACGGCGAA 357
 363 uGlyValThrArgSerAspArgAlaGlyGlyValGlyValGly.... 378
 358 GGGCTGCTGTATCATACGCCGACATCGGACATGATTTGGGCGGAGC 407
 378 378
 408 CTACATACGACGACAGCTTCGCTTCACCTGACCGCATGTACAGCGCG 457
 379 ...SerGlyProAlaGlyAsp.....Al 385

458 CGAAATGACGATAGACAAATCATCGAGCGGCGAGTCCGCGC 507
 385 aspcysArgSerAspProProGluArgAlaIaGlyArgGly.....G 400
 508 AAGCGAAACACCGCCGCCGATACAGAGGCTCAACAAATCATCA 557
 400 LUGLAlaAspTrpGlyAspAlaHisLeuArgArgGln..... 412
 558 GGGCTGCGCGCGCGGAGGAGAACCATTCGCGCGAGTCCCTT 607
 413AlaAlaGlyArgAlaGlnHisTrpSerLeuGlnProAlaAl 427
 608 CTCCGACAGAGAGCGCGCG...CGTGGCGGATTTTTCGCAACCT 654
 427 aspalaglyArgGlyAlaIleArgAlaAspAlaCysLeuArgGlnGly 444
 655 GCATACACCATGACACATGCG.....GGCAAAAT 683
 444 yspGlnHisHisHisGlyCysHisHisLeuSerAlaArgArgAspLeu 460
 684 GGCACACGTCGCAAG 698
 461 GlyGlnArgGlyArg 465

seq_name: Swissprot.40:FIIL-HUMAN

seq_documentation_block: PRT; 416 AA.

ID FIIL_HUMAN STANDARD; PRT; 416 AA.
 AC P20930;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Filaggrin precursor (Fragment).
 GN FIG.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN (1)
 RP MEDLINE=89296901; PubMed=2740331;
 RA McKinley-Grant L.J., Idler W.W., Bernstein I.A., Parry D.A.D.,
 RA Canizazaro L., Croce C.M., Huebner K., Lessin S.R., Steinert P.M.;
 RT "Characterization of a cDNA clone encoding human filaggrin and
 RT localization of the gene to chromosome region 1q21.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:4848-4852(1989).
 RN (2)
 RP MEDLINE=96374388; PubMed=8780679;
 RA Senshu T., Kan S., Ogawa H., Manabe M., Asaga H.;
 RT "Preferential delamination of keratin K1 and filaggrin during the
 RT terminal differentiation of human epidermis.";
 RL Biochem. Biophys. Res. Commun. 225:712-719(1996).
 CC -1- FUNCTION: AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES
 CC DISULFIDE-BOND FORMATION AMONG THE INTERMEDIATE FILAMENTS DURING
 CC TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.
 CC -1- PTM: FILAGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE,
 CC HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY "TANDEM COPIES
 CC OF 324 AA, WHICH ARE NOT SEPARATED BY "LARGE LINKER". THE
 CC PRECURSOR IS DELETED AS KERATOHYALIN GRANULES. DURING TERMINAL
 CC DIFFERENTIATION IT IS DEPHOSPHORYLATED & PROTEOLYTICALLY CLEAVED.
 CC -1- PTM: Undergoes delamination of some arginine residues
 CC (citrullination).
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: M24355; AAA52454.1; -

DR PIR: A32947, A32947.
 DR MW: 135940;
 DR InterPro: IPR003303; Filaggrin.
 DR PRINTS: PR00487; FILAGRIN.
 KW Phosphorylation; Citrullination; Developmental protein.
 FT NON_TER
 SQ SEQUENCE 416 AA: 44105 MW: DEEA3218BA043F32 CRC64;

alignment_scores:
 Quality: 100.00 Length: 226
 Ratio: 0.917 Gaps: 8
 Percent Similarity: 48.230 Percent Identity: 20.354

alignment_block:

US-09-303-518d-571 x FILA_HUMAN

Align seg 1/1 to: FILA_HUMAN from: 1 to: 416

```

159 CCGCGCGCGCATCTGCGC.....CAATATCGCGCAGCGGGTTTGA 199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
169 ProArgThrSerArgAsnTrpGlySerSerPheSerGlnAspSerAsp 185
200 ACCCGCAGCAGCAGCAGCGGTCAACCGTTTGGGAAAGCGCAAAATGC 249
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
185 rGlnGlnHisSerGlnAspSerGlnArgTrpSerGlySerAla..... 199
250 GGTTCGAACTTGCCCGCGCTTTTCAAAAACCGGAGACATCGAAAC 299
200 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
300 AATGTTCAAAACGGGTACACGCTGGCAACAGTCAGCAGCGCTTTGACA 349
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206 SerAlaGlnGlnGlnLeuArgAspGlySerArgHisProArgSerHis 222
350 AGCGGAAAGGCTGCTCTTATCATACGCCGCA.....380
222 nGluAspArgAlaGlyHisGlnHisSerAlaAspSerSerArgGlnSer 239
381 .....CATCGCAGCTACGATTGGG.....CGAGCGCTACT 413
239 lYThrArgHisThrGlnThrSerSerGlyGlnAlaAlaSerSerHis 255
414 CAGCCAGCAGCTTCCGTTCCACCTGACCCGCTAGACCGCCGAAAA 463
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
256 GlnGlnAlaArgSerSerAlaGlyAspArgHisGlnSerHisGlnGln 272
464 TCAAGCGATAGACAAATCATGCAAGC.....GGCAGGGTGGC 504
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
272 nSerAlaAspSerSerArgHisSerGlyLleGlnHisGlnAlaSer 289
505 GCGAAAGCAAAACCGCCGCCACCG.....530
289 eAlaValArgAspSerGlnHisArgGlyTrpSerGlySerGlnAlaSer 305
531 .....CATCAAGGGGTCAAAATCATCAAGCCCTGGCG 568
306 AspArgnGlnGlnHisSerGlnAspSerSerPThrGlnSerValSerAla 322
569 CGGCGGAGCAACCATCATCTGCCCGCAGCAGCTCCCTCCCGCAGAA 618
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
322 sGlyGlnAlaGlySerHisGlnGlnSerHisGlnGlnSerAlaGly 339
619 GCGCGCGCGCTGTC.....GCGGATTTTTCGCAAAACCTGC 656
|||:|||||:|||||:|||||:|||||:|||||:|||||:
339 rGSerGlyGlnThrSerGlyHisSerGlySerPheLeuTrpGlnValSer 355
657 ATACACCAT.....GACACTGG 673
356 ThrHisGlnGlnSerGlnSerSerHisGlyTrpThrGlyProSerThr 372
674 CGGCAAAATTCGCAACGCTCAAAAGCGT 701
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

372 gGlyArgGlnGlySerArgHisGlnGln 381

seq_name: SwissProt_40:CCAB_RABIT

seq_documentation_block:

ID CCAB_RABIT STANDARD; PRT; 2339 AA.

AC 005152;

DT 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DE 15-JUL-1999 (Rel. 38, Last annotation update)

DE Voltage-dependent N-type calcium channel alpha-1B subunit (Calcium channel, L type, alpha-1 polypeptide isoform 5) (Brain calcium channel II) (BIII).

GN CACNA1B OR CACNA1A5 OR CACNA5.

OS Oryctolagus cuniculus (Rabbit).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

OX NCBI_TaxID=9986;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=93236885; PubMed=8386525;

RA Fujita Y., Mylnieff M., Dirksen R.T., Kim M.-S., Nidome T., Nakai J.,

RA Friedrich T., Iwabe N., Miyata T., Furutachi T., Furutama D.,

RA Mikoshiba K., Mori Y., Beam K.G.;

RT "Primary structure and functional expression of the omega-conotoxin-sensitive N-type calcium channel from rabbit brain.";

RL Neuron 10:585-598(1993).

-1- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDATE THE ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION, CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1B GIVES RISE TO "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND ARE BLOCKED BY OMEGA-CONOTOXIN-GVIA (OMEGA-CTX-GVIA) AND BY OMEGA-AGATOXIN-IIA (OMEGA-AGA-IIIA).

CC BY OMEGA-CONOTOXIN-GVIA (OMEGA-CTX-GVIA) AND BY OMEGA-AGATOXIN-IIA (OMEGA-AGA-IIIA). THEY ARE HOWEVER INSENSITIVE TO DIHYDROPYRIDINES (DHP), AND OMEGA-AGATOXIN-IVA (OMEGA-AGA-IVA).

CC CALCIUM CHANNELS CONTAINING ALPHA-1B SUBUNIT MAY PLAY A ROLE IN DIRECTED MIGRATION OF IMMATURE NEURONS.

CC -1- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- TISSUE SPECIFICITY: WIDESPREAD EXPRESSION THROUGHOUT THE BRAIN.

CC -1- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.

CC -1- PTM: PHOSPHORYLATED IN VITRO BY CAM-KINSE II, CARX, PKC AND GSK3 (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: D14157; BAA0302.1; -

DR InterPro: IPR002077; Ca_channel.

DR InterPro: IPR002111; Cal_channel_TrpL.

DR InterPro: IPR000636; Cation_chan_non_lig.

DR InterPro: IPR001682; Channel_pore_Ca_Na.


```

712 TTTTCTGCTGCGAACCCCTGCGCCGACGACGCTTCGTTGCACAT 761
    :: ||||| ||||| ||| |||
981 .....GluGluProAlaArgHisAlaAlaGlyAla 993
762 CCGCCCGTCCAGGCGGATTGAACGCGACACA 794
    ||||| |||||
994 ProProlGlnGlnThrAlaGlnLysAspLys 1004

seq_name: SwissProt_40:EVAL_HUMAN

seq_documentation_block:
ID   EVAL_HUMAN          STANDARD:      PRT:      592 AA.
AC   099502:
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Eyes absent homolog 1.
GN   EYAL1.
OS   Homo sapiens (Human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX   NCBI_TaxId=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Embryo;
RX   MEDLINE=98030537; PubMed=9361030;
RA   Abdelhak S., Kalatzis V., Hellig R., Compain S., Samson D.,
RA   Vincent C., Levi-Acobas F., Cruaud C., Le Merrier M., Mathieu M.,
RA   Koenig R., Vigneron J., Weissenbach J., Petit C., Weill D.,
RT   "Clustering of mutations responsible for branchio-oto-renal (BOR)
RT   syndrome in the eyes absent homologous region (EYHR) of EYAL1."
RT   Hum. Mol. Genet. 6:2247-2255(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM EYAL1).
RC   TISSUE=Embryo;
RX   MEDLINE=97172972; PubMed=9020840;
RA   Abdelhak S., Kalatzis V., Hellig R., Compain S., Samson D.,
RA   Vincent C., Weill D., Cruaud C., Sahly I., Leboyevic M.,
RA   Bihner-Glindcz M., Francis M., Lacombe D., Vigneron J.,
RA   Charachon R., Boyen K., Bededer P., van Regemorter N.,
RA   Weissenbach J., Petit C.;
RT   "A human homologue of the Drosophila eyes absent gene underlies
RT   branchio-oto-renal (BOR) syndrome and identifies a novel gene
RT   family."
RT   Nat. Genet. 15:157-164(1997).
RL   Nat. Genet. 15:157-164(1997).
CC   -1- FUNCTION: MAY BE REQUIRED FOR NORMAL DEVELOPMENT OF BRANCHIAL
CC   ARCHES, EAR AND KIDNEY.
CC   -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; EYAL1 (SHOWN HERE) AND EYAL1B;
CC   ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -1- TISSUE SPECIFICITY: IN THE EMBRYO, HIGHLY EXPRESSED IN KIDNEY WITH
CC   LOWER LEVELS IN BRAIN, WEAKLY EXPRESSED IN LUNG. IN THE ADULT,
CC   HIGHLY EXPRESSED IN HEART AND SKELETAL MUSCLE. WEAKLY EXPRESSED IN
CC   BRAIN AND LIVER. NO EXPRESSION IN EYE OR KIDNEY.
CC   -1- DISEASE: DEFECTS IN EYAL1 ARE THE CAUSE OF BRANCHIO-OTO-RENAL
CC   (BOR) SYNDROME, AN AUTOSOMAL DOMINANT DISORDER MANIFESTED BY
CC   CYSTS, LACRIMAL DUCT STENOSIS, HEARING LOSS, STRUCTURAL DEFECTS OF
CC   THE OUTER, MIDDLE, OR INNER EAR, AND RENAL DYSPLASIA. ASSOCIATED
CC   DEFECTS INCLUDE ASTHENIC HABITUS, LONG NARROW FACES, CONSTRICTED
CC   PALATE, DEEP OVERBITE, AND MYOPIA. HEARING LOSS MAY BE DUE TO
CC   MONDINI TYPE COCHLEAR DEFECT AND STAPES FIXATION.
CC   -1- SIMILARITY: BELONGS TO THE EYA FAMILY.
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL: Y10260; CA71309.1; -
DR   EMBL: AJ000097; CA03922.1; -

```

```

DR   EMBL: AJ000098; CA03923.1; -
DR   MIM: 601653; -
DR   MIM: 113650; -
DR   InterPro: IPR001454; Hydrolase.
DR   Pfam: PF00702; Hydrolase; 1.
KW   Developmental protein; Multigene family; Alternative splicing;
KW   Disease mutation.
FT   VARSPLIC 1 41
FT   VARIANT 487 487
FT   FT
FT   VARIANT 505 505
FT   FT
SQ   SEQUENCE 592 AA; 64593 MW; D62365F81EB692E2 CAC64;
    /FTID=VAR_005204.
    /FTID=VAR_005204.

alignment_scores:
Quality: 99.00 Length: 394
Ratio: 0.559 Gaps: 22
Percent Similarity: 44.924 Percent Identity: 20.051

alignment_block:
US-09-303-518D-571 x EVAL_HUMAN ..

Align seg 1/1 to: EVAL_HUMAN from: 1 to: 592

6 TCGTTTACAAATTCAGGCTGTTCCCTTGGCAACCGGCATGCACATCC 55
|||||: |||: |||||:
73 SerPheSerProArgProThrHisGlnPheSerProProGlnIleTyrPr 89
56 TGTGACCGCCCTGCTCA...ATGCTCTCTCCCTGCTGCTTCC 99
|: ||||| |||||
89 OSerAsnAlaProTyrProHisIleLeuProThrProSerSerGlnThrM 106
|||||: |||||: |||||: |||||:
100 TGTCTGCACACGCTGGGAACCGCTCGACATCGGGCTT 140
|||||: |||||: |||||: |||||: |||||:
106 etAlaIleTyr...GlyGlnThrGlnPheThrGlyMeGlnGlnAla 121
141 .....TTACCTTTAAAGGAACCGCGCGCATCTGCG 175
|||||: |||||: |||||: |||||:
122 ThrAlaTyrAlaThrTyrProGlnProGlnIleProTyrGly...Ile 137
176 CCATATGCGCA.....GGGCGTTTAAAC 201
|||||: |||||: |||||: |||||: |||||:
137 rSerTyrGlnAlaLeuTyrPalaGlyIleTyrThrGlnGlyLeuSerG 154
202 CCGGACACGACGAGTCAAGCGTTTTCGGAACGCGCAATG... 248
|||||: |||||: |||||: |||||: |||||:
154 InsSerGlnSerProGlnIleThrGlnPheLeuSerTyrGlnThrSerPhe 170
249 .....CGTTTGGAACTTGCCCGCGCTTTTCAAAAACCGGAG 289
|||: |||||: |||||: |||||: |||||:
171 SerThrProGlnProGlnGlnAlaProTyrSerTyrGlnMetGlnLys 187
290 ACAT.....CGAAACATGTTCAAGCGGTACAGCGCTGGAG 330
|||||: |||||: |||||: |||||: |||||:
187 rSerPheThrThrSerSerGlyIleTyrThrGlnYAsnAsnSerLeuThr 204
331 GTGACGAGCC..... 341
204 snSerSerGlyPheAsnSerSerGlnGlnAspTyrProSerTyrProSer 220
342 TTTGGACAAAGGCGAGGCTGCTGTATCATCAGCC.....GC 379
|||||: |||||: |||||: |||||: |||||:
221 PheGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 237
380 ACATGCGACGCTACGATTTGGGGGAGCGTACATCAGCGACGCTTCG 429
|||||: |||||: |||||: |||||: |||||:
237 ahIstYrMetThrSerSerAsnThrSerProThrThrProSerThrAsn 254
430 TT.....CCACGTGACCGCCATGTA...CAACGCGC 458
|||: |||||: |||||: |||||: |||||:
254 IahThrTyrGlnLeuGlnGlnProProSerGlyIleThrSerGlnAlaVal 270

```


804TGCCGCGGTTCACCCGATACCGAATATTCGAT 839
 1012 splyglubylsglubeargashsglnpfoargluProHISCSASP 1028
 840 ACCGCGTTTCGACGCGAGTATCGT 866
 1029 LeuGlutHrSerglyThrValThrVal 1037
 seq_name: SwissProt_40:SPR2_HUMAN
 seq_documentation_block:
 ID SPR2_HUMAN STANDARD; PRT; 221 AA.
 AC 001130;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Splicing factor, arginine/serine-rich 2 (splicing factor SC35) (SC-35)
 DE (Splicing component, 35 kDa) (PR264 protein).
 GN SFRS2
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 67-84.
 RX MEDLINE=92237694; PubMed=1373910;
 RA Fu X.-D., Maniatis T.;
 RT "Isolation of a complementary DNA that encodes the mammalian splicing
 factor SC35.";
 RL Science 256:535-538(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RX MEDLINE=92212859; PubMed=1557353;
 RA Vellard M., Sureau A., Soret J., Martinier C., Perbal B.;
 RT "A potential splicing factor is encoded by the opposite strand of the
 trans-spliced c-myc exon.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:2511-2515(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94134745; PubMed=8302870;
 RA Sureau A., Perbal B.;
 RT "Several mRNAs with variable 3' untranslated regions and different
 stability encode the human PR264/SC35 splicing factor.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:932-936(1994).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Strausberg R.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE=93101590; PubMed=1465383;
 RA Sureau A., Soret J., Vellard M., Crochet J., Perbal B.;
 RT "The PR264/c-myc connection: expression of a splicing factor
 modulated by a nuclear protooncogene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:11683-11687(1992).
 RN [6]
 RP SEQUENCE OF 67-83.
 RX MEDLINE=92249775; PubMed=1577277;
 RA Zahler A.M., Lane W.S., Stolk J.A., Roth M.B.;
 RT "SR proteins: a conserved family of pre-mRNA splicing factors.";
 RL Genes Dev. 6:837-847(1992).
 RN [7]
 RP PROTEIN INTERACTIONS IN SPLICEOSOME ASSEMBLY.
 RX MEDLINE=94084782; PubMed=8261509;
 RA Wu J.Y., Maniatis T.;
 RT "Specific interactions between proteins implicated in splice site
 selection and regulated alternative splicing.";
 RL Cell 75:1061-1070(1993).
 RN [8]
 RP BINDING TO UI-70K.

RX MEDLINE=94187841; PubMed=8139654;
 RA Kohz J.D., Jamison S.F., Will C.L., Zuo P., Luhrmann R.,
 RA Garcia-Blanco M.A., Manley J.L.,
 RT "Protein-protein interactions and 5'-splice-site recognition in
 RT mammalian mRNA precursors.";
 RL Nature 368:119-124(1994).
 RN [9]
 RP RNA BINDING SPECIFICITY.
 RX MEDLINE=95354672; PubMed=7543047;
 RA Tacke R., Manley J.L.,
 RT "The human splicing factors ASF/SF2 and SC35 possess distinct,
 RT functionally significant RNA binding specificities.";
 RL EMBO J. 14:3540-3551(1995).
 RN [10]
 RP SPECIFICITY FOR BETA-GLOBIN MRNA.
 RX MEDLINE=93368668; PubMed=8361546;
 RA Fu X.-D.;
 RT "Specific commitment of different pre-mRNAs to splicing by single SR
 RT proteins.";
 RL Nature 365:82-85(1993).
 CC -1- FUNCTION: NECESSARY FOR THE SPLICING OF PRE-MRNA. IT IS REQUIRED
 CC FOR FORMATION OF THE EARLIEST ATP-DEPENDENT SPLICING COMPLEX AND
 CC INTERACTS WITH SPLICEOSOMAL COMPONENTS BOUND TO BOTH THE 5' AND 3'
 CC SPLICING SITES DURING SPLICEOSOME ASSEMBLY. IT ALSO IS REQUIRED FOR
 CC ATP-DEPENDENT INTERACTIONS OF BOTH U1 AND U2 SNRNPs WITH PRE-
 CC MRNA. INTERACTS WITH OTHER SPLICEOSOMAL COMPONENTS, VIA THE RS
 CC DOMAINS, TO FORM A BRIDGE BETWEEN THE 5' AND 3' SPLICING SITE
 CC BINDING COMPONENTS, U1 SNRP AND U2AF. IN VITRO, BINDS SF2/ASF,
 CC UI-70K AND THE 35 KDA BUT NOT THE 65 KDA SUBUNIT OF U2AF. BINDS TO
 CC PURINE-RICH RNA SEQUENCES, EITHER 5'-AGSAGAGAGA-3' (S-C OR G) OR
 CC 5'-GTTGAGTA-3'. CAN BIND TO BETA-GLOBIN MRNA AND COMMIT IT TO THE
 CC SPLICING PATHWAY.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: DIFFERENT FORMS OF THE PROTEIN MAY BE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE RS
 CC DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M90104; AAA60306.1; -
 DR EMBL: X62447; CAA44307.1; -
 DR EMBL: X75755; CAA53383.1; -
 DR EMBL: BC000039; AAH00339.1; -
 DR EMBL: L03693; AAA60162.1; -
 DR PIR: S17328; S17328.
 DR PIR: A42634; A42634.
 DR HSP: P09651; IUP1.
 DR MIM: 600813; -
 DR InterPro: IPR000504; RRM.
 DR Pfam: PF00076; rrm; 1.
 DR SMART: SM00360; RRM; 1.
 DR PROSITE: PSS00102; RRM; 1.
 DR PROSITE: PSS00030; RRM_RNP_1; 1.
 KW Nuclear protein; RNA-binding; mRNA splicing; Alternative splicing;
 KW Phosphorylation.
 FT DOMAIN 14 92 RNA-BINDING (RRM).
 FT DOMAIN 111 116 GLY-RICH (HINGE REGION).
 FT DOMAIN 117 221 ARC/SER-RICH (RS DOMAIN).
 FT CONFLICT 38 R -> G (IN REF. 2 AND 3).
 FT SEQUENCE 221 AA; 25575 MW; 9D1B76BDB65701F5 CRC64;

alignment_scores: 98.50 Length: 222
 quality: 98.50

Ratio: 0.895 Gaps: 11
Percent Similarity: 49.550 Percent Identity: 24.324

Alignment block:

US-09-303-518D-571 x SFR2_HUMAN

Align seg 1/1 to: SFR2_HUMAN from: 1 to: 221

```

199 AACCCGACAGCAGCGTCAAAACGCTTTGGGAAAGGCAAAATG 248
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
26 SerProasp.....ThleuArgrValPheglutylrArgrArgrva 40
249 CGGTTGGACACTGCCCC..... 267
40 ldlrvalrValrThrleProArgrArgrValrThrlygluSerArgrglp 57
268 ..GCGTTT.....TTCAAAAACCGAGACATCGAAACATGTTCAAA 309
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
57 heAlarPheValArgrPheHisAsrPlsrArgrAsrAlaGlulAsrAlaMetAsp 73
310 GCGGTACACGCGCTGGGAACACGTGACAGCGCTTGGACAAAGGCGCAAG 359
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
74 AlarMetAsrPlgl.....AlaValleuAsp...GlyArgrglu 84
360 GCTGCTGTTCATCAGCGCGGACATCGGACATGTTGGCGGACGCT 409
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
85 LeuArgrValrGluMetAla.....ArgTyrlelArgr..... 94
410 ACATTCAGCAGCAGCTTCGCTTCACCTGACCGCCATGTTCAGACGCGCG 459
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
95 .....ProProAsrPserHis...HisSerArgrA 103
460 AAATCAAGAGATGACAAATATCATGACGCGGCGGCGGCAAA 509
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
103 rglrProrProArgrArgrTyrlelYrlyglYrlyglYrlyglArgrArgrSer 119
510 AGGCAAAACCGCGCCACCGCATACAGGGGTCAACAAATCATCAAG 559
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
120 ArgrSerProArgrArgrArgrArgrArgrArgrArgrArgrArgr 136
560 CCCTGCGCGCGGCGGCGGACCATCATCTGCGGACGCGCTCCCTCT 609
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
136 rArgrSerArgrSerArgrSerArgrSerArgrSerArgrSerArgr 153
610 CCGCAGGAAGCGCGCGCTGTGGGCGATTTTTCGGAACCTGCATA 659
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
153 hrArgrSerArgrSerArgr..... 158
660 CACCATGACATGCGCGCAAAATTTGGACACGTCAGGCGTGAAGAACCC 709
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
159 .....SerThrSerArgrSerArgrSerArgrSerArgrSerArgr 173
710 TGTGTTTTCGTCGACGCGCTGCGCGA...CGGACAGGCTGTGTTG 756
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
173 rSerSerValrSerArgrSerArgrSerArgrSerArgrSerArgr 190
757 CACATCCGCGCGCGCTCCAAAGGGAATTCAGCGCAACAAACCCAGATGC 806
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
190 rSerProArgrProArgrSer.....LysArgrgluSerLysSerArgrSer 204
807 CGCGGTGTTCACACG 821
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
205 ArgrSerLysSerPro 209
seq_name: SwissProt_40:VE2_HPV37
seq_documentation_block:
ID VE2_HPV37 STANDARD; PRT; 454 AA.
AC Q80903:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE Regulatory protein E2.

```

```

GN E2. Human papillomavirus type 37.
OS Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=37958;
RN [1]
RP SEQUENCE FROM N.A.
RA Delius H.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC IT BINDS TO THE E2E RESPONSE ELEMENT (5'-ACNNNNNGGT-3') PRESENT
CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING ON E2E'S POSITION
CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC BY SPERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC REPLICATION.
CC -!- SUBUNIT: BINDS DNA AS A DIMER.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U31786; AAU9446.1; -.
DR HSSP: P17383; IDHM.
DR InterPro: IPR001866; E2_N.
DR InterPro: IPR000427; Early2_C.
DR Pfam: PF00511; E2_C; 1.
DR Pfam: PF00508; E2_N; 1.
DR ProDom: PD00672; Early2_C; 1.
DR ProDom: PD00678; E2_N; 1.
KW Early protein; Transcription regulation; Activator; DNA-binding;
KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
SQ SEQUENCE 454 AA; 51285 MW; ACA45808A7EA17AA CRC64;

```

alignment_scores: Quality: 97.00 Length: 275
Ratio: 0.789 Gaps: 14
Percent Similarity: 44.727 Percent Identity: 24.000

alignment_block:

US-09-303-518D-571 x VE2_HPV37

Align seg 1/1 to: VE2_HPV37 from: 1 to: 454

```

101 GTCCTGCACAGCGCTG.....GAAACCGGCTTGGACATCTG 135
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
130 ValTyrThrAlaTrpSerPheleTyrTyrGlnThrValaAsrPrlrrr 146
136 GCGTTTACCTTTTAAAGAACACCGCGCGCATGTCGCCATATGCG 185
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
146 P.....AsnLysVal.GluGlyHisValAsrTyrTyrly 157
186 GCAGCGGGTTTGAACCCCGACAGCAGCGTCAAGCCGTTTTCGG 235
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
158 AlaTyrTyrPheglu.....G1 163
236 AACGCAAAATGCGGTTT.....GGAACCTTGGCCCCGGGTTT 273
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
163 yAsrPleuLysValrTyrTyrleGlnPhegluGlyAsrAlaAlaArgr...P 179
274 TTCAAAAACCGGAGACATCGAAACATGTTCAAGCGGTACA..... 317
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
179 heserLysThrGlyArgrTrpGluValHisValasnuLysAsrPrllePhe 195
318 .....CGGCTGGACACGTCAGCAGCGTTCGTCAGGCGGCAAG 358
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

```

196 AlaProvalThrSerSerProAlaAlaGlyGlyThrAspGlyAl 212
359 GCGTGGTTCATCAAGCCGACATCGCAGTACGA.....TTGGGC 402
212 aaIaSerValHisThrValSerGlySerProLeuAlaArgGlyPheSer 229
403 GAGCGGTACATCAGCAGCAGCTCCGTTCCACCTGACGCGCATGTACA 452
229 hrThrSerValSerThrArgGlySerThrProArgArgTyrArgArg 245
453 GCGCCGCAAAAT..... 464
246 LysAlaSerSerProThrThrAlaAlaArgGlyLysArgGlyGly 262
465 .....CAAGGATAGACAAATCATGACGCGGCGAG..... 497
262 aaSPrrhAlaThrArgArgSerArgSerThrSerArgGlyLysGlnAla 279
498 .....GGTGGCGCGCAAAAGCAAAACCGCGCCGCGCATACA 536
279 hrSerArgGlyGlyAspArgArgArgArgArgGlyArgSerArgSer 295
537 AGGGGTCAAAATCATCAAGCCTCGCGCGGCGGCGGACCATCA 586
296 ArgAspThrSerSer.....SerProAspArgGlyArgGlyArgse 310
587 TCCGCGCCGACGACGCTCTTCCGCGAGAG..... 620
310 rArgGlyGlyProGluThrArgSerGlnSerArgSerLeuSerArgSera 327
621 .....CGCGCGGTGTGGCGGATTTTGGGCAACCTGCATACACCAT 665
327 rGserArgSerArgSerArgGlySerSerArgGlyValAlaPro 343
666 GACACTGGCGGCAAA.....ATTGCG..... 686
344 AspaIaValGlyLysSerValArgThrValGlyArgAspHisSerGlyAr 360
687 .....ACAGCTCAAGCGCTGAAGACCCGTTTCTGCTGC 723
360 gLeuLysArgLeuLeuAspGlnAlaArgAspProProValIleValLeuA 377
724 GAAGCGCTGGCGGAGCAGCAAG 746
377 rGgLyAspaIaAsnLysLeuLys 384

seq_name: SwissProt_40:IE63_HSV2H

```

```

seq_documentation_block:
ID IE63_HSV2H STANDARD; PRT; 512 AA.
AC P28276;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcriptional regulator IE63 (VMM63) (ICP27).
GN UL54.
OS Herpes simplex virus (type 2 / strain HG52).
OC Viruses: dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBL_TaxID=10315;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92113549; PubMed=1662697;
RA McGeeoch D.J., Cunningham C., McIntyre G., Dolan A.;
RT "Comparative sequence analysis of the long repeat regions and
RT adjoining parts of the long unique regions in the genomes of herpes
RT simplex viruses types 1 and 2."
RT J. Gen. Virol. 72:3057-3075(1991).
RL [2]
RP SEQUENCE FROM N.A.
RA Dolan A.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN THE DOWN-REGULATION OF VIRAL IMMEDIATE-EARLY

```

```

CC GENES AND ACTS IN COMBINATION WITH ICP0 AND ICP4 AS AN ACTIVATOR
CC OF LATE GENES (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL54,
CC HSV-2 UL54, EBV-1 5, VZV 4, EBV BMLF1, HCMV UL69, AND HVS-1 57.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D10471; BAA01269.1; -.
CC DR PIR: J01498; WMBEXA.
CC DR Early protein, transcription regulation; DNA-binding.
CC KW SEQUENCE 512 AA; 54958 MW; 459651470A503BA7 CRC64;

```

```

alignment_scores:
Quality: 97.00 Length: 135
Ratio: 1.386 Gaps: 9
Percent Similarity: 51.852 Percent Identity: 31.111

```

```

US-09-303-518D-571 x IE63_HSV2H ..

```

```

Align seg 1/1 to: IE63_HSV2H from: 1 to: 512

```

```

440 CCGCATGTACA.....AGCGCGCAAAATC 465
|||||:|||||
98 ProProAlaThrThrGlyValThrSerArgLeuGlyThrArgArgSerAl 114
466 AAACGATAGCAAAATCATGACGCGGCGGCGGCGGCAAAAGC 515
114 a.SerProArgGlyProHisGlyGlyValAlaArgGlyLeuProPro 130
516 AACCGCGCCCAAGCGCATACAGGGGTCAAAACATCATCAAGCCCTGC 565
131 SerThrLysAlaProHisProArgGly.....GlyArgArgGlyArgArg 145
566 GCGCGGCGGAGCAACCATCATCT.....GCCGACACAGCTC 603
145 gArgGlyArgGlyArgGlyArgGlyArgGlyAlaAspSerThrProLysP 162
604 CTTCTCCGAGGAGGCGGCGGTGTGGCGGATTTTTCGCAAGC 653
162 ro.....ArgArgArgValSer.....ArgAsnAla 170
654 TGCAATACACATGACACTGGCGCAAAATTGCGACAGCTCAAGCGCTGA 703
171 .....HisAsnGlnGlyArgHisProAlaSerAlaArgThrAs 184
704 AAACCGTGTCTTCTGCTGCGAAGCCCTCCCGCAG.....ACAAAGC 747
184 rGlyPro....GlyAlaThrHisGlyGlyAlaArgArgGlyGlyGlnL 200
748 TTGCGTGTGACATCGCGCGCTCCCAAGGGGATTTGAAGCGCAAAAGC 797
200 euAspValSerGlyGlyProArgProArgGlyThrArgGlnAlaProPro 216
798 CCA 800
|||
217 Pro 217

seq_name: SwissProt_40:YDB4_SCHPO
seq_documentation_block:
ID YDB4_SCHPO STANDARD; PRT; 297 AA.
AC Q10357;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)

```

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 32.2 kDa protein C22E12.04 in chromosome I.
 GN SPAC22E12.04.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 NC NCB1_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; 270043; CAA93891.1; -
 CC DR HSSP; P04355; 4MT2.
 KW Hypothetical protein.
 SO SEQUENCE 297 AA; 32243 MW; 70C128A3BE2B2BA CRC64;

alignment_scores:

Quality: 96.50 Length: 173
 Ratio: 1.109 Gaps: 10
 Percent Similarity: 50.289 Percent Identity: 23.699

alignment_block:

US-09-303-518D-571 x YDB4_SCHPO ..

Align seg 1/1 to: YDB4_SCHPO from: 1 to: 297

```

109 AGCGTGGGAAACCGGCTCGACATCTGGCGGTTTACCTTTAAG..... 153
    ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
163 SerLeuProAsnTrpIleGlyHisCys.....PheValLeuLysCysVal 177
154 .....GAAGACCGCGCGCATGCTGCCAATATGCGGCGAGGGGGT 196
    ||||| ||| ||| ||| ||||| ||||| ||||| ||||| |||||
177 LaspAspSerAspSerAlaThrMetGlyTlleSerAlaGlySerAlaGlyL 194
197 TGAACCCGACAGCAGACGAGGTCAAGCGCTTTTGGGAAGCGCAAA 246
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
194 eugly.....GlnAsnThrLysGlnIleCysAlaCysThrGlyLys 207
247 TCGGGTTTGAACTTGGCCCCCGGTTTTCAAAAAACCGGAAGACATGCA 296
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
208 .SerLeuTrpThrGlnHisAlaGlnLeuLysSerValAsnGlnGlySer 224
297 AACAAATGTTCAAGCGGTACACGGCT.....GGGAACACGGTC 334
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
224 eTcYcYcSerLysLysAspSerSerProSerGlnLysProSerGlyCys 240
335 ACCAGGCTTTGGACAAGGCGGAGGCTGCTGTTTCATCAGCGGCACATC 384
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
241 SerGln.....GlnLysLysSerGlyCysSerLysLys..... 252
385 GGCACCTACGATTTGGCGGAGCGTACATCAAGCAGACTTCGTTCCA 434
252 .....
435 CCGTACCCCATGTACAGCCGCGGAATATCAAGCATAGACAATAATCA 484
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
253 .....ProSerGlyCysSerGlnGlnLys.....LysGly 262
485 TGC.....AGGCGGAGGAGGTGCGCGGCAAGGCAAAACCGCGCC 525
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
263 CysCysSerThrGlnLysThrSerGlyCysSerGlnGlnLys..... 276

```

526 ACCGGCATCAAGGGGTCAACAAATCATCAAGCGCTTGGCGGCGCA 575
 277
 576 GGCAACCATCATCTTGC 592
 283 LulysProSerGlyCys 288

seq_name: SwissProt_40:CRE1_NEUCR

seq_documentation_block:

ID CRE1_NEUCR STANDARD; PRT; 430 AA.

AC 059958;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE DNA-binding protein cre-1 (Carbon catabolite repressor).

GN CRE-1.

OS Neurospora crassa.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariatales; Sordariaceae; Neurospora.

NC NCB1_TaxID=5141;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=74-OR23-1A;

RA de la Serna I., Tyler B.M.;

RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: INVOLVED IN CARBON CATABOLITE REPRESSION. REPRESSIONS THE

CC IN THEIR PROMOTER (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).

CC -!- SIMILARITY: BELONGS TO THE CRE1/MIG GROUP OF C2H2-TYPE ZINC-

CC FINGERS PROTEINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF05464; AAC13555.1; -

DR HSSP; P07248; 1ABD.

DR InterPro: IPR000822; Znf-C2H2.

DR Pfam: PF00096; Zf-C2H2; 2.

DR PRINTS: PR00048; ZINC_FINGER.

DR SMART: SM00355; Znf_C2H2; 2.

DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 2.

DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 2.

KW DNA-binding; Transcription regulation; Repressor; Zinc-finger;

KW Nuclear protein; Metal-binding;

FT ZN_FING 78 100 C2H2-TYPE.

FT ZN_FING 106 130 C2H2-TYPE.

FT DOMAIN 36 45 POLY-GLN.

FT DOMAIN 142 146 POLY-GLN.

SO SEQUENCE 430 AA; 46954 MW; FDD304889A7D8085 CRC64.

alignment_scores:

Quality: 96.00 Length: 297
 Ratio: 0.857 Gaps: 12
 Percent Similarity: 37.710 Percent Identity: 22.222

alignment_block:

US-09-303-518D-571 x CRE1_NEUCR ..

Align seg 1/1 to: CRE1_NEUCR from: 1 to: 430

```

8 GTTACATTCAGGCGTGTTCCTTTTGGCAACGCCA..... 46
    ||||| ||| ||| ||||| ||||| ||||| ||||| |||||
150 ValHisAsnHisGlyLeuGlnProAspMetMetProProGlyProly 166

```

```

47 .TGACATCTGTTGACCGCCCTGCTCAATGCTCTGCTGCTGCT 95
166 salalaleargseralaprothralametserserprobsanvalserp 183
96 TTCCGCTGTCACAGCTGGGAAACCGGCTCGGACATCGGCTTTTAC 145
183 roPro.HisSerTySerProTyraInphelaproserglyleuanphr 199
146 TTTTAAAGGAAGACCGCGCCGATCGCCCAATATCGGACGCGGCT 195
199 oTy.....SerHisSerArgSerSerAlaIyser.... 209
196 TTGAACCCCGACAGCAGCGTCAAAACCGTTTTCGGAACCGCAAA 245
209 ..... 209
246 ATGCGGTTTGAACCTGCCCGCGTTTTCAAAAACGGAGACATCG 295
209 ..... 209
296 AAACATGTTCAAACCGCTACACGGCTGGAGAACATGACAGGCTTG 345
210 .....GlnSerGlyProAspIleSerLeuAlaIargAlaIag 223
346 GACAAAGGCGGAGG.....GCTGCTGTCATCACGCGGACATCGGAG 389
223 LglInValGlnIargAspGlyAlaAlaInshishPhedInProIarghe 239
390 CTACGATTTGGGCGGAGCTACATGACCGACGCTTC.....GT 430
240 GlnPhetGlyAsnThrLeuHisAlaIatThrAlaSerArgSngIle 256
431 TCCA.....CCTGACCGCCATGTATCAAGCCCGCAAAATCAAGCATA 474
256 uproGlyLeuGlnAlaIatYHisMeSerArgSerHisSerHisgluapH 273
475 GACAAATCATCGACGCGGAGGCTGCCGGAAGCAAAACCGCGCC 524
273 IsasprshistYgIy.....GlnSerTyArgHisAla 284
525 CACCGGCATACAAAGGGGTCAAAATCATCAAGCCCGCGCGGGGG 574
285 LysArgSerArgProAsnSerProAsnSerThrAlaProSerSerProth 301
575 AGGCAACATCATCTCGCCGACGAGCTCTTCGCGAGGAGGCGGC 624
301 rPheserHisasprSerLeuSerProthPro..... 311
625 GCGGTGTGGGCGGATTTTTCGGAACCTGCATACCATGACACTGSC 674
311 ..... 311
675 GGCAGAAATGGCACAGCTCAAGGCGTGAACCCGTGTTTCTGCTGCG 724
312 .....AspHisThrPro.....LeuAl 317
725 AACGCTGCGCGACGAGCAAGGCTTCGTGTCACATCG..... 764
317 atHrProAlaHisSerProArgLeuArg....ProHisProGlyLeuGluL 333
765 .CCCCGTCAGAGGAATTTGAAGCGCAACAAACCCAGAGATCGCGCT 812
333 euProProPhetArgAsnLeuSerLeuGlyInGlnHisThrThrProAla 349
813 GTTCAACCGCAATACGCAATATGATGATGACGCGTTTTC 851
350 LeuThrProLeuGlnProAlaLeuAspGlyGlnPheser 362
seq_name: SwissProt_40:NONA_DROME
seq_documentation_block:
ID NONA_DROME STANDARD; PRT; 700 AA.

```

```

AC 004047;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE No-on-Transient A protein.
GN NONA.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90262721; PubMed=2344408;
RA Jones K.R., Rubin G.M.;
RT "Molecular analysis of no-on-Transient A, a gene required for normal
RT vision in Drosophila."
RL Neuron 4:711-723(1990).
CC -!- FUNCTION: REQUIRED FOR NORMAL VISION IN DROSOPHILA.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; I (SHOWN HERE) AND II: ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: M33496; AAA03214.1; -.
DR EMBL: M33496; AAA03215.1; -.
DR HSSP: P19339; 1SXL.
DR FLYBASE: FBgn0004227; nona.
DR InterPro: IPR000504; RRM.
DR Pfam: PF00076; rrm; 2.
DR SMART: SM00360; RRM; 2.
DR PROSITE: PS50102; RRM; 2.
DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
KW RNA-binding; Repeat; Vision; Alternative splicing.
FT DOMAIN 302 374 RNA-BINDING (RRM) 1.
FT DOMAIN 376 457 RNA-BINDING (RRM) 2.
FT DOMAIN 73 76 POLY-GLY.
FT DOMAIN 254 264 POLY-GLY.
FT DOMAIN 619 626 POLY-GLY.
FT DOMAIN 648 652 POLY-ASN.
FT VARSPLIC 666 700 DSFAFEFGVNNMNOGNGNGNGGNNVPGRRRF -> VC
FT PPKRYPTKYSVTNRSEVLEIQLSMIPMKL (IN
FT ISOFORM II).
SQ SEQUENCE 700 AA; 76967 MW; 732DB77FC5DFD47 CRC64;

alignment_scores:
Quality: 95.50 Length: 221
Ratio: 0.910 Gaps: 10
Percent Similarity: 47.511 Percent Identity: 23.529

alignment_block:
US-09-303-518D-571 x NONA_DROME ..

Align seg 1/1 to: NONA_DROME from: 1 to: 700

183 GCGGAGCGGGGTTGAACCCGACAGCAGCGTCAAGCGGTTTTC 232
|||||
77 AlavalGlyGly.....ProsnInshslnysasnph 87
|||||
223 CGGAACCGCAAAATGCGGTTTGAACCTGCCCGCGGTTTTCAAAAA 282
|||||
87 eGlyAsnshslnysGlyGlyPhe.....ValG 96
|||||
283 CCGAGACATCGAACAATGTTCAAGCGGTACAGCGTGGGACAC... 329
|||||

```



```
96  lYAsnArGAsnAlYAsnAsnAsnArgAlaGlyAsnGlnAsnArgThrPhe 112
330 .....CGTGCAGCAGGCTTGG 346
113  ProGlyAsnAsnSerAsnGlnLysProAsnAsnGlnThrSerLysAl 129
347  ACAGGGCGCAGAGGCTGCTTCATCAGCGCCGACATCGCAGCTACGAT 396
129  aAspGlyProAsnAlaLeuAlaLysAsn..... 138
397  TTGGGGGAGCGCTACATCAGCCAGCAGCTCCGTTCCACTGACCGCCAT 446
139 .....AsnGlnProAlaThrAlaAla..... 145
447  GTACAGCCGCGCAAAATCAAGCGATAGACAAATCATCGACGGCGGCA 496
146 .....AlaGlyGlnAsnGlnAlaAsn...GlnAsnAlaAsnLysGlyGln 159
497  GGGTGGCGGCGCAAGGCAAAACCCGCCACCGGCATACAGGGGTCAAA 546
159  nAsnGlnArgGlnGlnAsnGlnAsnGlnAsnGlnValHisGlyGlnG 176
547  CAATCATCAAGGCCCTGCGCGC.....GGG 572
176  lYAsnGlnGlyLysProGlyAsnGlnGlyAlaGlyAsnGlnGlyGly 192
573  CGAGGCAACCATCATCTGCCCCGACGATCCCTCTCCGCGAGGAAGCG 622
193  GlnGlyAsnGlnGly.....GlyAlaGlyAsnGln 202
623  GCGGCGTGTGGGCGGATTTTTCGGCAAACTGATACACCATGACACTG 672
202  nGlyAsnGlyGlnGlyPheArgGlyArgAsnAlaGlyAsnAsnGlnGlyG 219
673  GCGCAAAATTGGC.....ACAGTCAAGGCGTGAAAAACCTGTTT 716
219  LysGlyPheSerGlyGlyProGlnAsnGlnGlnArgAspAsn..... 232
717  CTGCTGGAGACGGCTGCGCAGCAGACAGGCTTCGTGTGCACATCCGCC 766
233 .....ArgAsnArgSerGlyPyr 238
767  CCGTCCAAGGGGA 779
238  oArgProGlyGly 242
```

